

This page Is Inserted by IFW Operations
And is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of
The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

XX WPI: 2000-452368/39.
 DR P-SDB: AAY90265.
 XX Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,
 PT prevention and treatment of Chlamydia infection in mammals -
 XX
 PS Claim 2; Fig 1; 81bp; English.
 XX This sequence encodes the Chlamydia pneumoniae ATP/ADP translocase
 CC of the invention. The protein, DNA encoding it, or a vaccine containing
 CC the DNA or protein, are useful for diagnosing, preventing or treating
 CC Chlamydia infection. The sequences can also be used in a method for
 CC the detection of Chlamydia infection. Primers or probes derived from the
 CC DNA sequence are useful in diagnostic tests for detecting Chlamydia
 CC infection.
 XX
 SQ Sequence 1637 BP; 431 A; 327 C; 325 G; 554 T; 0 other:
 Query Match 100.0%; Score 1637; DB 21; Length 1637;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 721 TACGCTTTTGATGCGTATGACATATGATCTGACCTGTTCTTATGCGCAGTTACTGCT 780
 Oy 781 GGATCAATAGACGATGTTGACCGATCCGCTTCATATATCCAGAAATGCAAAAG 840
 Db 781 GGATCAATAGACGATGTTGACCGATCCGCTTCATATATCCAGAAATGCAAAAG 840
 Oy 841 GGAAGAAAGGCTAAACCTTAATGAATATGAAGATAGCTTCCTATCTTATAGAT 900
 Db 841 GGAAGAAAGGCTAAACCTTAATGAATATGAAGATAGCTTCCTATCTTATAGAT 900
 Oy 901 CTCCTTATATCTTTTATATACCTCTCTGTTGTTATGCTATGATTTGCTTAA 960
 Db 901 CTCCTTATATCTTTTATATACCTCTCTGTTGTTATGCTATGATTTGCTTAA 960
 Oy 961 TCGAAGTGACTTGGAAAAAGTACGTAACATGCAATATCCATATGATGATATAGTG 1020
 Db 961 TCGAAGTGACTTGGAAAAAGTACGTAACATGCAATATCCATATGATGATATAGTG 1020
 Oy 1021 AGTTCATGGGGAACCTCTCTTCTGACCTGGCTGATATCCGTAATGCTATTTG 1080
 Db 1021 AGTTCATGGGGAACCTCTCTTCTGACCTGGCTGATATCCGTAATGCTATTTG 1080
 Oy 1081 TTGCTGGTACGCTATCTGTAATTTGGATGTTAACTGAGCCCTAGTACTCTGTCA 1140
 Db 1081 TTGCTGGTACGCTATCTGTAATTTGGATGTTAACTGAGCCCTAGTACTCTGTCA 1140
 Oy 1141 TGGTTCTCCACAGATATCGCTTCTTGGCTCTTGTATCTTATAGAACCAACCTCTG 1200
 Db 1141 TGGTTCTCCACAGATATCGCTTCTTGGCTCTTGTATCTTATAGAACCAACCTCTG 1200
 Oy 1201 GCGTGTGCGTATGTTGGTACCAACTCTCTCATGCTAGCTGTGTTGCGAGCTATAC 1260
 Db 1201 GCGTGTGCGTATGTTGGTACCAACTCTCTCATGCTAGCTGTGTTGCGAGCTATAC 1260
 Oy 1261 AGAATATCTTGGAAATCCACAATAATAGCGCTCTTGTGACTCAATGAAGAAATGGCCT 1320
 Db 1261 AGAATATCTTGGAAATCCACAATAATAGCGCTCTTGTGACTCAATGAAGAAATGGCCT 1320
 Oy 1321 ATATCCCTCTTGACACAGACCAAAAGTCAAAAGGTAAGGCTGATATGATAGTCCG 1380
 Db 1321 ATATCCCTCTTGACACAGACCAAAAGTCAAAAGGTAAGGCTGATATGATAGTCCG 1380
 Oy 1381 CCGCCTTCGGAATTCAGAGAGAGCTTTAATCCACAAGGTTGCTGTTATCTGTGAA 1440
 Db 1381 CCGCCTTCGGAATTCAGAGAGAGCTTTAATCCACAAGGTTGCTGTTATCTGTGAA 1440
 Oy 1441 GTATTGAGCTATGACCCCTTATCTGACAGGATCTCTTTTCATATGCTATTTGGT 1500
 Db 1441 GTATTGAGCTATGACCCCTTATCTGACAGGATCTCTTTTCATATGCTATTTGGT 1500
 Oy 1501 TGGTTCTGCAACTAATGTTAAACAACTATCTTAGCGAGCTGCTTTAAAGAACAG 1560
 Db 1501 TGGTTCTGCAACTAATGTTAAACAACTATCTTAGCGAGCTGCTTTAAAGAACAG 1560
 Oy 1561 AAGTGGCTCAAGAAATTCAGCTCCGCTTCTTCATATAGAGTGTGCTCTTACTGTT 1620
 Db 1561 AAGTGGCTCAAGAAATTCAGCTCCGCTTCTTCATATAGAGTGTGCTCTTACTGTT 1620
 Oy 1621 GATCCCTACCTGCTTT 1637
 Db 1621 GATCCCTACCTGCTTT 1637

RESULT 2
 AAX91990
 ID AAX91990 standard; DNA: 1230025 BP.
 XX AAX91990:
 AC
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 XX

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.
 OS Chlamydia pneumoniae.

XX MO9927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (GSEST) GENSET.

XX Griffals R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Claim 1; Page 291-611; 1912pp; English.

XX The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 99.8%; Score 1633.8; DB 20; Length 1230025;

Best Local Similarity 99.9%; Pred. NO. 0;

Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAATATAAACTATCAGATAGAAATTAAGTATTCAGAGGCTAAATATGACAAAA 60
 Db 404320 GAAATATAAACTATCAGATAGAAATTAAGTATTCAGAGGCTAAATATGACAAAA 404379
 QY 61 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTCTGTGGCCGATACATCTCAGAGC 120
 Db 404380 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTCTGTGGCCGATACATCTCAGAGC 404439
 QY 121 TAAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGTATTCATTTAACTATACGGTGT 180
 Db 404440 TAAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGTATTCATTTAACTATACGGTGT 404499
 QY 181 TACGGATACAAAAGACACTCTTATTTGGGAGCTCTGTTCTGTGTCAGAGGCAATATAC 240
 Db 404500 TACGGATACAAAAGACACTCTTATTTGGGAGCTCTGTTCTGTGTCAGAGGCAATATAC 404559
 QY 241 CTTTCATCAAGTTTGGGCTGTGTGTCGCCCTGTGCTATATCTTTATGCTATTTATGCAA 300
 Db 404560 CTTTCATCAAGTTTGGGCTGTGTGTCGCCCTGTGCTATATCTTTATGCTATTTATGCAA 404619
 QY 301 AGCTAAGTAAATATTTAAGTAAAGAGGCTTATTTATGACAGTGGGAAGGCCCTTTTAA 360
 Db 404620 AGCTAAGTAAATATTTAAGTAAAGAGGCTTATTTATGACAGTGGGAAGGCCCTTTTAA 404679
 QY 361 TTTTCTTGGCCCTGTTCGCCAGTGAATTTATCCGCTACGCGANTTTTACATCCCTACAG 420
 Db 404680 TTTTCTTGGCCCTGTTCGCCAGTGAATTTATCCGCTACGCGANTTTTACATCCCTACAG 404739
 QY 421 AATTGCGACCGTTTACAGGCGATCCACCTCCAGAGATGTTAGAGACTCGTGGCACTCT 480
 Db 404740 AATTGCGACCGTTTACAGGCGATCCACCTCCAGAGATGTTAGAGACTCGTGGCACTCT 404799

QY 481 TAAGAACTGACATTTGCTGCAATTTATAGTACTGTGAACTATGAGGGAACGTCATGC 540
 Db 404800 TAAGAACTGACATTTGCTGCAATTTATAGTACTGTGAACTATGAGGGAACGTCATGC 404859
 QY 541 TATCTAATGTTCTGGGGATTTGCTAATGAATTAACAAAAATCCAGCAAGCAAGCGTT 600
 Db 404860 TATCTAATGTTCTGGGGATTTGCTAATGAATTAACAAAAATCCAGCAAGCAAGCGTT 404919
 QY 601 TCTACGCTCTTTGCGTATCGAGCTAATATTTCTTTACCTAGCTTCTGCTGCAATTC 660
 Db 404920 TCTACGCTCTTTGCGTATCGAGCTAATATTTCTTTACCTAGCTTCTGCTGCAATTC 404979
 QY 661 TTTGGGCTTCAAAAGTTGAGAGCTTCGTTTGAAGGTTAGATCCTTTGGGAATTCCT 720
 Db 404980 TTTGGGCTTCAAAAGTTGAGAGCTTCGTTTGAAGGTTAGATCCTTTGGGAATTCCT 405039
 QY 721 TACGCTCTTTGATGGCTATGACTATTTATGATGCACTTGTCTTATGCGCACTTACTGT 780
 Db 405040 TACGCTCTTTGATGGCTATGACTATTTATGATGCACTTGTCTTATGCGCACTTACTGT 405099
 QY 781 GGATCAATTAAGACGCTATTTGACCGATCCTCGCTTATATTCAGAGAAATGCAAAAG 840
 Db 405100 GGATCAATTAAGACGCTATTTGACCGATCCTCGCTTATATTCAGAGAAATGCAAAAG 405159
 QY 841 GAAAAAAGGCTCTAACCTTAATGAATGAATGAAGATAGCTTCCCTATCTGATAGAT 900
 Db 405160 GAAAAAAGGCTCTAACCTTAATGAATGAATGAAGATAGCTTCCCTATCTGATAGAT 405219
 QY 901 CTCTTATATCTTTTATTAATCTCTTGTGTTATGCTTATGCTTATGCTTATTAATTA 960
 Db 405220 CTCTTATATCTTTTATTAATCTCTTGTGTTATGCTTATGCTTATGCTTATTAATTA 405279
 QY 961 TCGAAGTACTTGGAAAAAGTACGCTGAACCTGCAATATCTAATATGAAATGACTATAGTG 1020
 Db 405280 TCGAAGTACTTGGAAAAAGTACGCTGAACCTGCAATATCTAATATGAAATGACTATAGTG 405339
 QY 1021 AGTTATGAGGGAATTCCTCTCTGAGCTGGGCTGATTCGCTATTCATGATGATTTG 1080
 Db 405340 AGTTATGAGGGAATTCCTCTCTGAGCTGGGCTGATTCGCTATTCATGATGATTTG 405399
 QY 1081 TTGCTGTAAGCTCAATTCGTAATTTGATGGTAACTGAGAGCCCTAGCTACTCCTGTCA 1140
 Db 405400 TTGCTGTAAGCTCAATTCGTAATTTGATGGTAACTGAGAGCCCTAGCTACTCCTGTCA 405459
 QY 1141 TGGTCTCTTAACAGGATTCGTTTCTTCTGCTCTTGTATCTTAAGAAACCAAGCTTCTG 1200
 Db 405460 TGGTCTCTTAACAGGATTCGTTTCTTCTGCTCTTGTATCTTAAGAAACCAAGCTTCTG 405519
 QY 1201 GCGTGGTGGCTATGTTGGTACAACTCGCTCATGCTAGCTGCTGTTGGAGGCTATATAC 1260
 Db 405520 GCGTGGTGGCTATGTTGGTACAACTCGCTCATGCTAGCTGCTGTTGGAGGCTATATAC 405579
 QY 1261 AGAATATTTCTTTCGAATTCACAAATACGCTCTCTTGGACTCAACTAAAGAAATGGCCT 1320
 Db 405580 AGAATATTTCTTTCGAATTCACAAATACGCTCTCTTGGACTCAACTAAAGAAATGGCCT 405639
 QY 1321 ATATCCCTCTTTCGACCAAGCAAAAAAGTCAAGGCTGCTATTTATGATGATGCTCGC 1380
 Db 405640 ATATCCCTCTTTCGACCAAGCAAAAAAGTCAAGGCTGCTATTTATGATGATGCTCGC 405699
 QY 1381 CCGGCTTGGAAAAATCAGAGAGGCTTAAATCCCAAGAGTTGCTGTTATTCGTGAAA 1440
 Db 405700 CCGGCTTGGAAAAATCAGAGAGGCTTAAATCCCAAGAGTTGCTGTTATTCGTGAAA 405759
 QY 1441 GTATTGAGCTATGAGCCCTTATCTTTCAGAGTATCTTCTTTCATCAATTCCTATTTGGT 1500
 Db 405760 GTATTGAGCTATGAGCCCTTATCTTTCAGAGTATCTTCTTTCATCAATTCCTATTTGGT 405819
 QY 1501 TGGTTTTCGCAACTTAAGTTAAACAACATTTCTTATGCGCAGTCTGCTTTAAAGAACAG 1560
 Db 405820 TGGTTTTCGCAACTTAAGTTAAACAACATTTCTTATGCGCAGTCTGCTTTAAAGAACAG 405879

QY 1561 AAGTGGCTCAAGAGATTACACTGCTCTTCTCATAGAGTTGCTTCTTACTCTTGT 1620
|||||
Db 405880 AAGTGGCTCAAGAGATTACACTGCTCTCTCTCATAGAGTTGCTTCTTACTCTTGT 405939
QY 1621 GATCCCTACCTGCTTTT 1637
|||||
Db 405940 GATCCCTACCTGCTTTT 405956

RESULT 3

AAC81914/c

ID AAC81914 standard; DNA: 273254 BP.

AAC81914:

DT 27-FEB-2001 (first entry)

DE Chlamydia pneumoniae genome DNA.

KW Genome: diagnosis; vaccine; ds.

OS Chlamydia pneumoniae.

PN M0200027994-A2.

PD 18-MAY-2000.

PF 12-NOV-1999; 99MO-US26923.

PR 12-NOV-1998; 98US-0108279.

PR 08-APR-1999; 99US-0128606.

PA (REGC) UNIV CALIFORNIA.

PI Stephens R, Mitchell W, Kalman S, Davis R;

DR WPI: 2000-376516/32.

PT Isolated nucleic acid for use in diagnostic and analytical methods

PT encodes genomic sequence of Chlamydia pneumoniae -

PS Claim 2; Page 128-320; 320pp; English.

XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC pneumoniae protein (P1), given in the specification. The isolated nucleic
CC acid is useful for diagnostic and analytical methods, such as,
CC hybridization-based assays or amplification-based assays. The protein may
CC be used for diagnostic purposes, for their enzymatic or structural
CC activity, or as a vaccine. The invention also describes (1) a probe
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC cassette comprising N1 under the transcriptional regulation of a
CC transcriptional initiation region functional in an expression host, and a
CC transcriptional termination region; (4) a cell comprising an expression
CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell as a result of induction of the expression
CC cassette into the host cell, and the cellular progeny of the host cell;
CC (5) a method for producing a P1 comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other
CC proteins; (6) a purified polypeptide composition comprising at least 50
CC weight % of P1; and (7) a monoclonal antibody binding specifically to the
CC peptide of (6).

SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Query Match 99.74; Score 1632.2; DB 21; Length 273254;

Best Local Similarity 99.86; Pref. No. 0;

Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATATAAAACATCATGATAGAAAATAAAGTATTTACAGAGGTAATATGACAAAA 60
|||||
Db 212470 GAATATAAAACATCATGATAGAAAATAAAGTATTTACAGAGGTAATATGACAAAA 212411

QY 61 CCGAAGAAAACCTTTGGAAAATTTGGCTCTTCTGTGGCCGATACATACGAGC 120
|||||
Db 212410 CCGAAGAAAACCTTTGGAAAATTTGGCTCTTCTGTGGCCGATACATACGAGC 212351
QY 121 TAAAGAAAGTCTGCCAATGTTCCTAATGTCTCTGTATTAACATTAACGCGT 180
|||||
Db 212350 TAAAGAAAGTCTGCCAATGTTCCTAATGTCTCTGTATTAACATTAACGCGT 212291
QY 181 TACGGATACAAAAGACACTTATGTGAGAGCCGCTTGTGTCAGAGCAATAC 240
|||||
Db 212290 TACGGATACAAAAGACACTTATGTGAGAGCCGCTTGTGTCAGAGCAATAC 212231
QY 241 CTTTCATCAAGTTTGGCTGTGTCCCTGTGCTATTAATCTTAATGCTTATTAAGCAA 300
|||||
Db 212230 CTTTCATCAAGTTTGGCTGTGTCCCTGTGCTATTAATCTTAATGCTTATTAAGCAA 212171
QY 301 AGCTAAGTAATATTTAGTAAGCAGGCTTATTTATGCAAGTGGAAAGCCCTTTTAA 360
|||||
Db 212170 AGCTAAGTAATATTTAGTAAGCAGGCTTATTTATGCAAGTGGAAAGCCCTTTTAA 212111
QY 361 TTTTCTTTGGCCCTGTCCGACTGTAATTTATCCGCTACGCGATGTTTACATCCTACAG 420
|||||
Db 212110 TTTTCTTTGGCCCTGTCCGACTGTAATTTATCCGCTACGCGATGTTTACATCCTACAG 212051
QY 421 AATTTGCTGACCGTTTACAGGCAATCCATCCCTCAGAGATTGCTAGACTGTTGCCATCT 480
|||||
Db 212050 AATTTGCTGACCGTTTACAGGCAATCCATCCCTCAGAGATTGCTAGACTGTTGCCATCT 211991
QY 481 TAAAGAACTGGACATTTTGCTCATTTATGTACTTGCTGCAACTATGGGAAGCCTCATGC 540
|||||
Db 211990 TAAAGAACTGGACATTTTGCTCATTTATGTACTTGCTGCAACTATGGGAAGCCTCATGC 211931
QY 541 TATCTTAATGTCTTGAGGATTTGCTAATGAATTAACAAAATCCAGAAAGCAAGCTT 600
|||||
Db 211930 TATCTTAATGTCTTGAGGATTTGCTAATGAATTAACAAAATCCAGAAAGCAAGCTT 211871
QY 601 TCTACGCTCTTTGCGATGCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATG 660
|||||
Db 211870 TCTACGCTCTTTGCGATGCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATG 211811
QY 661 TTTGGGCTCAAGTTAGAGACTTCGTTCTGGAAGGTGATCCTTGGGAATTTCTT 720
|||||
Db 211810 TTTGGGCTCAAGTTAGAGACTTCGTTCTGGAAGGTGATCCTTGGGAATTTCTT 211751
QY 721 TACGCTTTTATGATGCTATGCTATTTGATGCACTTGTCTATGCGCAGTACTG 780
|||||
Db 211750 TACGCTTTTATGATGCTATGCTATTTGATGCACTTGTCTATGCGCAGTACTG 211691
QY 781 GGATCAATAGACGATATGACGATCCCTGCTTATATCAGAAAGAAATGCAAAAG 840
|||||
Db 211690 GGATCAATAGACGATATGACGATCCCTGCTTATATCAGAAAGAAATGCAAAAG 211631
QY 841 GGAAAAAAGGTGTAACCTTAATGAATGAAGATAGCTTCCCTATCTTGCTAGAT 900
|||||
Db 211630 GGAAAAAAGGTGTAACCTTAATGAATGAAGATAGCTTCCCTATCTTGCTAGAT 211571
QY 901 CTCCTTATATCTTTTATTAAGCTCTTGGTATATGCTATGATTTGATTAACCTTAA 960
|||||
Db 211570 CTCCTTATATCTTTTATTAAGCTCTTGGTATATGCTATGATTTGATTAACCTTAA 211511
QY 961 TCGAAGTCACTTGGAAAAGTCAGCTGAACCTGCAATATCTTAATATGATGATATG 1020
|||||
Db 211510 TCGAAGTCACTTGGAAAAGTCAGCTGAACCTGCAATATCTTAATATGATGATATG 211451
QY 1021 AGTTCATGGGGAAGCTTCCCTGCTGAGCTGGCGATATCCGTAATCATGCTATTTG 1080
|||||
Db 211450 AGTTCATGGGGAAGCTTCCCTGCTGAGCTGGCGATATCCGTAATCATGCTATTTG 211391
QY 1081 TTTGCTGAACGTCATTCGTAATTTGATGCTTAACCTGAGCCCTAGTCACTCTGTCA 1140
|||||
Db 211390 TTTGCTGAACGTCATTCGTAATTTGATGCTTAACCTGAGCCCTAGTCACTCTGTCA 211331
QY 1141 TGGTTCTCTAACAGGATGCTTTCTTCGCTCTTGTATCTTAAACCAAGCTTCTG 1200

```

|||||
Db 211330 TGGTCTCTACAGATGCTTTCTGCTCTTCTGATTCCTTAGAACCAAGCTTCG 211271
OY 1201 GGCTGTGCTATGTCGCTGACACCTCTCTACTAGCTGTGTTGCGAGCTATAC 1260
Db 211270 GGCTGTGCTATGTCGCTGACACCTCTCTACTAGCTGTGTTGCGAGCTATAC 211211
OY 1261 AGAATATCTTTCGAATCCACCAAAATAGCTCTTTGACCTAAAGAAATGGCCT 1320
Db 211210 AGAATATCTTTCGAATCCACCAAAATAGCTCTTTGACCTAAAGAAATGGCCT 211151
OY 1321 AATATCCCTTTCGAACCAAGCAAAAGTCAAGAGTGTCTATTTGATTTAGTTGCG 1380
Db 211150 AATATCCCTTTCGAACCAAGCAAAAGTCAAGAGTGTCTATTTGATTTAGTTGCG 211091
OY 1381 CCGGCTTCGGAATCAGAGAGCTTTATATCACAAGGTTGCTGCTATCTGTCGAA 1440
Db 211090 CCGGCTTCGGAATCAGAGAGCTTTATATCACAAGGTTGCTGCTATCTGTCGAA 211031
OY 1441 GATTGAGCTATGACCCCTTATCTGAGTGATTCCTTTTCATCATTTGCTATTTGGT 1500
Db 211030 GATTGAGCTATGACCCCTTATCTGAGTGATTCCTTTTCATCATTTGCTATTTGGT 210971
OY 1501 TGGTTTCTCACTAAGATTAAACAAATATTTCTAGCGAGTCTGCTTTAAAGAACAG 1560
Db 210970 TGGTTTCTCACTAAGATTAAACAAATATTTCTAGCGAGTCTGCTTTAAAGAACAG 210911
OY 1561 AAGTGGCTCAAGAGATTAGCTCTGCTTCTTCATAGAGTTGCTTCTTACTCTTGT 1620
Db 210910 AAGTGGCTCAAGAGATTAGCTCTGCTTCTTCATAGAGTTGCTTCTTACTCTTGT 210851
OY 1621 GATCCCTACCTGCTTTT 1637
Db 210850 GATCCCTACCTGCTTTT 210834

```

RESULT 4
AAZ01425
ID AAZ01425 standard; DNA; 1038602 Bp.

XX AAZ01425;

XX 07-OCT-1999 (first entry)

XX Complete genome sequence of Chlamydia trachomatis.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KM Paratrachoma; inclusion conjunctivitis; genital disease; perithelitis;
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX Bartholinitis; pneumonia; venereal lymphogranulomatosis; ss.

OS Chlamydia trachomatis.

XX W09928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-1B01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97PR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST) GENSET.

XX Griffais R.

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Claim 1; Page 373-656; 1755bp; English.

CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AY36754-137949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g., eye diseases such as
CC conjunctivitis, genital diseases such as nongonococcal urethritis,
CC epididymitis, cervicitis, salpingitis, perithelitis, Bartholinitis;
CC pneumonia in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.

XX Sequence 1038602 Bp; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 53.5%; Score 876; DB 20; Length 1038602;
Best Local Similarity 73.0%; Pred. No. 16-229;
Matches 1155; Conservative 0; Mismatches 421; Indels 7; Gaps 2;

```

OY 1 GAATATAAAAACATATCAGAAATAGAAAATTAACATTTTCAAGAGGT-AAATATGACAAA 59
Db 392356 GATCAACAAAAGAAACCTTAACATTAACCTTTTCAGAGGGTAAATATGACTCAA 392415
OY 60 ACCGAGAAAAACCTTTTGGAAAAATGCGCTCTTCTGCGCCGATACATACTACAGAG 119
Db 392416 ACCGCGAAAAACCTTTTGGAAAAATGCGCTCTTCTTGGCCGATACATGATGAG 392475
OY 120 CTAAAGAAAGTCTGCAATGTTCTTAATGTTCTTCTGATTAATACATTAATACAGTG 179
Db 392476 CTGAAGAAAGTCTGCAATGTTCTTAATGTTCTTCTGATTAATACATTAATACAG 392535
OY 180 TTACGGATACAAAACACACTTATTTGGAGGCTCTGTTCTGTCGACAGGCAATA 239
Db 392536 TTGAGAGATACAAAAGATCTTATGCTTACGACACCGGAGCTGAGACAGGCAAT 392595
OY 240 CCTTCATCAAGCTTTTGGCTTGTGTCCTGCTGCTATTTATCTTATGCTATTTATGCA 299
Db 392596 CCTTCATTAAGTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 392655
OY 300 AAGCTAAGTAATATTTTAAAGTAAGAGCGCTTATTTTATGACAGTGGGAAAGCCCTTTT 359
Db 392656 AAGCTAAGCAATATTTTGAACAAACAGGCTCTTCTTTCGACAGTGTCTACATCGT 392715
OY 360 ATTTCTTTGCCCTGTTCCCGACTTAATTTATCCGCTACGAGATTTTATCTCTA 419
Db 392716 GATTTCTTGGCACTGTTCCCTGCTGATCTACCCCTGCGCTATTTCTTCAACCCAMA 392775
OY 420 GAATTTGCTGACCGTTTACAGGCACTACCTCCAGAGTTGCTAGACGCTGCTGATC 479
Db 392776 GCTTTGCTGATACCTTACAAATCATCTTCTCCGATTTATGAGATTTCAATGCGATG 392835
OY 480 TTAAAGAACTGACATTTGCTGATTTTATGATTTGCTGCACTATGAGGAAAGCGTCAG 539
Db 392836 CTAGCAACCTGACATTTGCTGATTTGCTGATTTGCTGCACTATGAGGAAAGCGT 392895
OY 540 CTATCTCTAATGTTCTGAGGATTTGCTAATGAATTAACAAAATCCACAGCAAGCGT 599
Db 392896 CTCTCTTTATGTTCTGAGGATTTGCTAATGAATTAACAAAATGCAAGGTAACGT 392955
OY 600 TTCTACGCTTTTGGATTTGAGAGCTATATTTTCTAGCTCTGCTGCTGCTGCTGCT 659
Db 392956 TTCTACGCTTTTGGAGATGAGAGCTAATGATCTCTTCTTAAATTTCTGCTGCTGCT 393015
OY 660 GTTTGGCTTCAAGATTGAGAGCTTCTCTTCTGAGAGTGTAGATCTTGGGAAATTTCT 719
Db 393016 ATCTGCTTTCTAAATTTACGCTTACCTTGGGAAAGCGTTGACCAAGGAGTCAAC 393075
OY 720 TTACGCTTTTATGCTATGATTTGATCTGCACTGTTCTTATGCGCCAGTTACTG 779
Db 393076 CTCTATTTCTTAATGCTATGCTTCTGCTGCTGCTTATGCTGCTGCTGCTGCTG 393135
OY 780 TGGATCAATAAGAGCTATGACCGATCTGCTCTTATTAATCCAGAAAGTGAAGAG 839
|||||

```

Db 393136 TGGATGAACCCGTCACGTGCTTACAGATTCCTACATCTCGACAGAGCTCAAAGCT 393195
Qy 840 GGGAAAAAGGTGCTAAACCTAAATGAATATGAAGATAGCTTCCTCATCTTGATAGA 899
Db 393196 AAGAAA-----TCTAAACCTAAGATGACATGGGGAAGAGCTTCTTATCTGTTAAGA 393249
Qy 900 TCTCCTTATATCTTTTATTAACCTCTTGGTTATGCTTATGCTTATGCTTAACTTAA 959
Db 393250 TCTCCTTATATCTTCTTCTACCTCTTCTAGTATTTGCTACAGGATCGATCTTAACCTC 393309
Qy 960 ATCGAAGTACCTGGAAAAAGGCAACCTACATATCTCAATATGAATGATACATATAGT 1019
Db 393310 GTTGAAGTACCTGGAAAAAGGCAACCTCAATATGCAATCTCAATCAATGATACACG 393369
Qy 1020 GAGTTCATGCGGAACTTCTCTTCTGAGTGGCGTAGTATCCGCTACTTATCATGCTATTT 1079
Db 393370 GCATTTATGGGGAACCTCTCTTCTGAGAGAGAGTGTATCTGATTTGCTATGCTCTTC 393429
Qy 1080 GTTGGTGTACAGCTATTCGTAAATTTGGATGTTAACTGAGAGCCCTACTCTCTCTTC 1139
Db 393430 ATCGGTGGTAAAGCTATTCGTGATTCGGTGTAACTGAGAGCTTGTGTTACACCTTATC 393489
Qy 1140 ATGGTTCCTCTACAGTATGCTTCTGCTCTGTTATCTTAAAGAAACCAAGCTTCT 1199
Db 393490 ATGGTTCCTCTACAGTATGCTTCTGCTCTGTTATCTTAAAGATCATGCTTAC 393549
Qy 1200 GGGCTGTGCTATGTTGGGTACAACTCTCATGCTAGCTGTGTTGCTCGAGCTATA 1259
Db 393550 GGTGTTAGTGTGCTGTGGGAAACACACCTAATGCTGAGCTGTGTTGCTCGAGCTAT 393609
Qy 1260 CAGATATCTCTTGAATCCCAAAATACGCTCTTGTGACTCAACTAAACAAATGGGC 1319
Db 393610 CAAATATCTCTTCAAAATCAATCAATATGCTCTTGTGACTCAACTAAAGAGCTGGCT 393669
Qy 1320 TATATCTCTTGTGACCAAGAGCAAAAGTCAAGCTGATGATGATGATGATGCTGCC 1379
Db 393670 TACATCCCATCTGATCAAGAGCAAAAGTCAAGGAAAGCTGATGATGATGCTGCC 393729
Qy 1380 GCCCGCTTGGCAAAATCAAGAGAGCTTAACTCAACAAGGTTGCTGCTATGCTGGA 1439
Db 393730 GATATGATGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393789
Qy 1440 AGTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1499
Db 393790 AGCATGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 393849
Qy 1500 TTGCTTCTGCACTAAGTAAACAACTATTTAGCGCAGTGTGCTTAAAGACAA 1559
Db 393850 TTGACATCTGCACTAAGTAAACAACTATTTAGCGCAGTGTGCTTAAAGACAA 393909
Qy 1560 GAAGTGGCTCAAGAAATTCAGC 1582
Db 393910 GAATGAGNAGAAAGCTGACGACG 393932

RESULT 5
AAf61374
ID AAF61374 standard; cDNA: 1896 BP.
XX
AC AAF61374:
XX
DT 25-MAY-2001 (first entry)
XX
XX Potato chloroplast ATP/ADP translocator cDNA.
XX
DE ATP/ADP translocator; chloroplast; plant; adenosine triphosphate;
KM adenosine diphosphate; animal feed; energy; biosynthesis; potato;
KM nutritional value; ss.
XX
OS Solanum tuberosum.
XX
PN WO200120009-A1.

PD 22-MAR-2001.
XX
XX
PF 05-AUG-2000; 2000WO-EP07625.
XX
PR 15-SEP-1999; 99DE-1044212.
XX
XX (BADI) BASF AG.
XX
PI Reinold A, Geigenberger PL, Neuhaus H, Graeve-Kampfenkel K;
PI Moehlmann T, Tjeden J;
XX
XX WPI: 2001-244814/25.
XX
XX
XX Transformed plant with an altered content of essential amino acids and
PT having improved nutritional quality, for use as an animal feed,
PT comprises a modification in the ATP/ADP translocator gene -
PS
PS Example 4; Fig 2; 32pp; German.
CC
CC This invention describes a novel transformed plant (A), and its
CC descendants, in which the regulatory sequences and/or the gene copy
CC number of an ATP/ADP (adenosine tri-/di-phosphate) translocator gene
CC (I) has been altered so that, compared with the wild-type plant, it
CC has a different content of one or more amino acids (aa). The invention
CC also describes (1) (I), for use in plants, having an approximately 1.7
CC kb sequence (S1), given in the specification, from Arabidopsis thaliana,
CC EMBL 249227; (2) a genetic construct (GC) containing the (1) and operably
CC linked regulatory sequences; (3) a vector containing the (1) or GC; and
CC (4) seeds, tissues, cells and replicative materials of (A). (A) are
CC useful for human or animal feeding, also (including their cells, tissues
CC and extracts) in agriculture, the animal feed and pharmaceutical
CC industries, and the health service. Altering the activity of (1)
CC increases the amount of ATP in chloroplasts and the amount of energy
CC available for biosynthesis. Specifically (A) have increased contents of
CC one or more essential aa, so are of greater nutritional value than
CC wild-type plants.
XX
XX
SQ Sequence 1896 BP: 473 A; 369 C; 445 G; 609 T; 0 other:
Query Match 23.2%; Score 380.2; DB 22; Length 1896;
Best Local Similarity 55.1%; Pred. No. 2.6e-94;
Matches 804; Conservative 0; Mismatches 628; Indels 27; Gaps 2;
Qy 74 TTTTGGAAATTTGGCGCTTTCTTGTGCGCATATCACTACAGCTAAGAAATTTCT 133
Db 285 TGTGGAAGAAAGCAACCTAAGTTATGAGGATGAACTTGATGACCTTAAGAAATATAT 344
Qy 134 GCCATGTTCTAATGTTCTCTGATATACATTAACATACGATGTAACCCGATACAA 193
Db 345 ACCACTGGGCGATGTTCTTTGATTTCTGTTAATTAATCAATCCCTTAAGGATACAA 404
Qy 194 AGACACTCTTATTTGGAGCTCCGCTGTTGCGAGAGGCAATACCTTTCATCAAGTT 253
Db 405 GGAATGTTGTTGTTAAGAGCTAAAGAGGTCAGAGTGTGAGATTTTCCCTTTGGAAGAC 464
Qy 254 TTGCTTGTGCTCCCTGCTATTAATCTTATGCTTAATTAAGCAAGCTAAGTATAT 313
Db 465 TTGGGTGAATTTGGCTATGCTATGATGATGATGATGATGATGATGATGATGATGATG 524
Qy 314 TTTAAGTAAAGAGGCTTATTTATGAGTGGGAAGCCCTTTTAATTTCTTGGCCCT 373
Db 525 GTTGTCANAAGAGGCTCTTTTATTAATGCTTATTAATGCTTATGCTTATGCTTATGCT 584
Qy 374 GTTCCGACATTAATTTATCGCTACGAGATTTTATACCTACAGAAATTTGCTGACCG 433
Db 585 GTTGGTTTGTGTTTGTATCTCTTGAAGATTTACTTTACCCCTACAGGTTTGTGATAA 644
Qy 434 TTTAAGAGGCAATCTACCTCAGATGCTAGAGCTGTTGCCATCTTAAGAAATGAGAC 493
Db 645 GCTTCTCAATACCTGTGCTCAAGATTTCTTGACCAATTCCTATTCGAGAGATCTGGAG 704
Qy 494 ATTGCTGATTTTATTAATGCTGATGAGCTAATGGGAAGCTCATGCTATGATATGTT 553

Db 705 TTCTGCTTTGTTCTAATGATGAGCTTGAGCTTTGGGAGATGTCGTTCAGTACTCTT 764
Oy 554 CTGGGATTTTGCATTAATGAATTTACAAAATCCAGAGCAAAAGCTTTCTAGCCTCTTT 613
Db 765 TTGGGATTTTGCATTAATGATGATGAGCTTGAGAGCTTAAGATTTCTATCTTTGTT 824
Oy 614 CGGATTCGAGAGCTTAATTAATTTCTTTACTAGCTTCTGCTGCAATTTGGGCTTCAAA 673
Db 825 TGGACTTGGAGGAATGTTGCTCTTATTTTCTGTGTCGACAGTAAGTACTTTTCTAG 884
Oy 674 GTTGAAGACTTCCGTTTCTGAAAGTGTAGATCTTGGGAAATTTCTTACGCTTTTGTAT 733
Db 885 CTTGAGAAAGCTTTTATGAGTCTGAGAGTGTATGTTGGGTATCTCCCTGAAAGATGAT 944
Oy 734 GGCTATGATATTTGATCTGAGAGCTTGTCTTATGAGCAATTTCTGCTGATCAATTAAGA 793
Db 945 GAGTATTTGTTGATGATGAGGAGGCAATCTGTTCTTTACTGTGGGTGAATGAGAAA 1004
Oy 794 CGTATTTGACCGATCTCGCTCTATTAATCCAGAGAAATGCAAAAAGGGGAAAAAGCTGC 853
Db 1005 TGTGTCTCCCACTCGTAGCA-----GAAGAGAGAGT 1040
Oy 854 TAACTTAATTAATGAATATGAAGATAGCTTCTATATCTGATAGATCTCTTATATTC 913
Db 1041 AAAACCTTAACATGACCAATGAGAGCTTGAAGTTCTTGCTCTTCAAAATATATACAG 1100
Oy 914 TTTATTAATCTCTGTTATGCTATGCTATGCTATTTGATTTGATTAATGCAAGTCTTG 973
Db 1101 GATCTCTGCACTGAGTGGTGTAGCAATATGAGCAATATCAACCTTTGAAAGTTTCAAG 1160
Oy 974 GAAAGTCACTGAAAGCTCAATATCTATATGATGATGATGATGATGATGATGATGATG 1033
Db 1161 GAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1220
Oy 1034 GTTTCCTCTGAGTGGCTGATATCTGATATCTGATATCTGATATCTGATATCTGATAT 1093
Db 1221 CTCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1277
Oy 1094 CATTCGTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1153
Db 1278 TTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1337
Oy 1154 AGCTATGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1213
Db 1338 CGGAGTGGTCTTCT 1397
Oy 1214 GTTTCGTAACACT 1273
Db 1398 GTTTCGTAATGATCT 1457
Oy 1274 GAAATCCCAAAATAGCT 1333
Db 1458 TAGAGTGCAGAAATAGTATGTTGTTGATGATGATGATGATGATGATGATGATGATG 1517
Oy 1334 CCAAGAGCAAAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1393
Db 1518 TAGAGCAACCAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1577
Oy 1394 ATCAGAGAGAGCTTTATCAACAAAGTTGCTCTGTTATCTGGAAGTATGAGAGCTAT 1453
Db 1578 GTCTGAGAGAGCTTTGATCAACAGTTGATGATGATGATGATGATGATGATGATGAT 1637
Oy 1454 GAGCCCTATCTTGGAGTATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1513
Db 1638 GACACCTTACCTTGGGAGTGTCTCTTATGATATGTTCTTGAATGTTGGAGACGACCA 1697
Oy 1514 TAACTTAACAACTATTC 1532
Db 1698 GTCTTTGATGAGAGTTC 1716

RESULT 6
AAC36046

ID AAC36046 standard; DNA; 2146 bp.
AC
XX AAC36046;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12350.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.

CC sinusitis. It is also linked to diseases other than respiratory tract
 CC infections, being associated with atherosclerosis and asthma.
 CC Npt2cp nucleotides may be used for the recombinant production of the
 CC protein, as genetic vaccines, in the construction of vaccine vectors
 CC such as poxviruses, and in the construction of attenuated Chlamydia
 CC strains which can over-express an Npt2cp nucleotide, or which can
 CC express it in a non-toxic, mutated form. Npt2cp polypeptides can be used
 CC in vaccine compositions, and may also be used as diagnostic reagents for
 CC detecting the presence of anti-Chlamydia antibodies in a sample.
 XX

Sequence 1823 Bp; 450 A; 359 C; 396 G; 618 T; 0 other:

Query Match 18.4%; Score 301.4; DB 22; Length 1823;
 Best Local Similarity 50.3%; Pred. No. 1.3e-72;

Matches 814; Conservative 0; Mismatches 776; Indels 27; Gaps 2;

0Y 2 AATATAAACTATCAGATATGAAATATAAAGTATTCAGAGGTAATATGACAAAAC 61
 Db 52 AACTAAATGCTATAGCTTTATACGCCATTAAGTTGGAGAGGTTATCATCATC 111
 0Y 62 CGAATAAAACCTTTGGAAATTCGCTCTTCTGGCCGATACATACACAGAGCT 121
 Db 112 AGAAGTGAACCTTTTCAAGGCTCGGGCATATCTTGTCTATTATATAATCAGAAAT 171
 0Y 122 AAGAAGATTCGCCAATGTCCTAATGTTCTCTGTAATTAATTAATTAATTAATTAAT 181
 Db 172 TTCTAGATTTGTTCCACATTTCTACTAGCGTTTTCGTGGCTTTAACTACTGCTGCT 231
 0Y 182 ACGCATACAAAAGACACTCTTATTTGTGGAGCTCCTGTTCTGGTGCAGAGCAATACC 241
 Db 232 GAAAAACATGAAAGATCTGCTGCTATGTCGTTTCAGATGCTGGGGCAGAGATATCC 291
 0Y 242 TTTTCATCATGTTTGGCTTTGTCGCCCTGCTATTAATCTTATAGCTTATTAATGCAAA 301
 Db 292 CTTCCTTAGGTTTGGGAATGTCCCGGAGCTGTATGTTACTATGTTATGGGTG 351
 0Y 302 GCTAAGTATATTTTAAAGTAAGCAGCCTTATTTATGAGTGGGAGAGCCCTTTTAAAT 361
 Db 352 GTTAGGAGTGGGTATCTCGGGATATCCGCTTTTATATGCTTCAATGCGCATCTTG 411
 0Y 362 TTTCTTGGCCCTGTCGCCGACTGTATTTTATCCGCTAGCCGATGTTTATACATCACA 421
 Db 412 TTTTTCCTTCGTTGCTGATCATTTATTCCTAGGAGATACCTGATCTCAACAC 471
 0Y 422 ATTTCGTAACCTTTACAGGCCATCTACTCCAGAGATTGCTAGAGACGTGTTCCATCT 481
 Db 472 TCTGCTATATAATTAACAAGAGCTCTTCCTCAAGAGACTCGTGTATTAATGATGCT 531
 0Y 482 AAGAAGTGAACATTTGCTGATTTTATGTAATGTAATGTAATGTAATGTAATGTAATG 541
 Db 532 CCGTACTAGGAGTATACAGATTTATTAAGTATGTAATGTAATGTAATGTAATGTAATG 591
 0Y 542 ATCTCTAATGTTTGGGATTTGCTAATGAATTAACAAAATTCACGAAGAAGCGTTT 601
 Db 592 TTTGATGTTGTTGGGACTAGGCAATCAGATTAATTAATTAATTAATTAATTAATTAAT 651
 0Y 602 CTACGCTCTTTTGGTATGCGAGCTAATTAATTTCTTACAGTCTGTCGTCGCAATGCT 661
 Db 652 TTACGCTCTTTATCAATACAGATTAATTAATCTCTCTCAATATGCGAGGGAATCTCCTA 711
 0Y 662 TTGGCTTCAAAAGTTGAGAGCTTCGTTTCTGAAGGTGATGCTTGGGGAATTTCTTT 721
 Db 712 TTGGATGGGGAACAAACATTTGTTGCGTACCTTTTGCATGTAATCTTGGCACTGCT 771
 0Y 722 ACGTCTTTGATGCTATGACTATTTGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 781
 Db 772 AATGCTCAACTGACATGCTGATGCTGTTCTGTTTAAATTAATGATGCTGCTATTAAG 831
 0Y 782 GATCAATTAAGAACTATGACGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 817
 Db 832 GCGGATTAATCACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891
 0Y 818 TAATTCAGAGAAATGCAAAAAGGGGAAAAAGGTGCTAATCAATTAATGAATATGAAGA 877

Db 892 ACAGGAGACGCTACTGCTAATCTAAGAGAAAAAGAAAAACCTAAAGCAAGCTAGAAA 951
 0Y 878 TAGCTTCCCTTATCTTGAATAGATTCCTTATATTTCTTTATTAATCTCTGTTATTCG 937
 Db 952 CCTTTCTTACACCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1011
 0Y 938 CTATGATTTTGCATTAATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 997
 Db 1012 CTATTAATTTGGATTCATCTATTCGAAGTCTTTTGGAGAGATCAAGATTAAGCAATTTA 1071
 0Y 998 TCCATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1057
 Db 1072 CAGTTCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1131
 0Y 1058 ATCCGACTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1117
 Db 1132 TTTCTGATTAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191
 0Y 1118 TGGAGCCCTAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1177
 Db 1192 CGGTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1251
 0Y 1178 TATCTTTAGAAACCAAGCT---TCTGGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1234
 Db 1252 TTTTGTGCAAAAAGAGACATCTATTTTGGGGAGTCTTGGATGACACCTGTGCG 1311
 0Y 1235 GCTAGCTGTTGTTGCGAGCTATACAGAAATTTCTTTGGAATCCACAAATTAAGCTCT 1294
 Db 1312 TCTAGCTGCTGAGCTGAGGAGTCAAAATGCTTATCCAGGAGCAAAATTAAGCTCT 1371
 0Y 1295 CTTCAGCTCACTAAAGAAATGGCTATATCCCTCTGACCAAGAGCAAAATTAAGCT 1354
 Db 1372 CTTTGATCAACCAAGAAATGGCTTATCCACTTCTCCAGAGGATTAATTAAGCT 1431
 0Y 1355 TAAGCTGCTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1414
 Db 1432 GAAAGCCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1491
 0Y 1415 ACAAGTTGCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1474
 Db 1492 CCAAGGCGCTGTTGATTTCTCTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
 0Y 1475 TCTCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1534
 Db 1552 TCTCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1611
 0Y 1535 AGCGAGTCTGCTTAAAGAACAGAAAGTGCCTCAAGAAATTCAGCTCGCTGCTT 1591
 Db 1612 CTCTAGAGCTGCTGATGCTGTAAGCAACCTTGAACCAACCTTAAGAACCTTCTCTT 1668

RESULT 9
 ABL74862
 ID ABL74862 standard; cDNA; 284 Bp.
 XX
 AC ABL74862;
 DT 14-MAY-2002 (first entry)
 DE
 XX Corn tassal-derived polynucleotide (cdps) seq ID NO:4236.
 XX
 KW Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPS;
 KW inheritance; characterstic; growth; development; disease resistance;
 KW environmental adaptability; quality; yield; molecular marker;
 KW multigene trait; plant breeding; corn tassal; gene; ss.
 OS
 XX Zea mays.
 XX
 XX US2001051335-A1.
 XX
 XX 13-DEC-2001.
 PD
 XX

PF	16-APR-1999;	99US-0294093.
XX		
PR	21-APR-1998;	98US-082567P.
XX		
PA	(LALG/) LALGUDI R V.	
PA	(ITOL/) ITO L Y.	
PA	(SHER/) SHERMAN B K.	
XX		
PI	lalgudi RV, Ito LY, Sherman BK:	
XX		
DR	WPI: 2002-163647/21.	
XX		
PT	Novel purified corn tassel-derived polynucleotide useful for	
PT	determining altered gene expression, to recover regulatory elements and	
PT	to follow inheritance of desirable characteristics through hybrid	
PT	breeding programs -	
XX		
PS	Claim 1; SEQ ID 4236; 201pp; English.	
XX		
CC	The present sequence describes a purified corn tassel-derived	
CC	polynucleotide sequence (cdps) comprising a nucleic acid sequence	
CC	selected from those given in ABL70627 to ABL76833. The cdps sequences	
CC	encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)	
CC	can be used for determining altered gene expression, to recover	
CC	regulatory elements and to follow inheritance of desirable	
CC	characteristics through hybrid breeding programs. (I) are also useful	
CC	in the evaluation, and alteration of desired characteristics associated	
CC	with growth and development, disease resistance, environmental	
CC	adaptability, quality and yield, and as molecular markers for studying	
CC	inheritance of multigenic traits in a plant breeding program. (I) can be	
CC	used to produce a tassel-specific profile of gene transcription, a	
CC	transcript image, to clone regulatory elements for use in transformation	
CC	vectors, to express a polypeptide, to identify, isolate or extend	
CC	identical or related corn tassel nucleic acid sequences from DNA	
CC	libraries, in nucleic acid hybridisation or amplification technologies,	
CC	as query sequences to determine homology of known sequences, as probe	
CC	for use in Southern or Northern hybridisation, and to identify the	
CC	presence of and/or to determine the degree of similarity between two	
CC	(or more) nucleic acid sequences.	
XX		
SQ	Sequence 284 BP: 66 A; 55 C; 75 G; 86 T; 2 other:	
	Query Match	5.9% Score 96.8; DB 24; Length 284;
	Best Local Similarity	61.3%; Pred. No. 1.4e-16;
	Matches 155; Conservative	0; Mismatches 98; Indels 0; Gaps
OY	1177 TTATCTTTAGAACAACCAAGCTTCGTGGCGTGCCTGATGTTGCGTAACACTCCTTCATGCG	1236
Db	1 TTTTGTTGTGGTAGAGCATTTGACTCCCTTTATTGACCAGAAATTTGGGATGACGCTTTCGTTG	60
OY	1237 TAGCTGTGCTGTGGGAGCTATACAGAAATATCTTTTGGAAATCCACAAAATFAGCTTCT	1296
Db	61 CGGAGCTGTATGTTGGGAGCAATGAGAAATTTTCAGTAGAGTGCAAAAAATACAGTCTGT	120
OY	1297 TTGACTCACTAAGAAATAAGGCTATATCCCTCTTGAGCAAGAACAAAAGTCAAAGCTA	1356
Db	121 TTGATCCCTTCGNMAAGATGSCATACATCTCTTTGGCATGAGGATATTAAGAGTGAAGGTA	180
OY	1357 AGCGTCTATTGATGTAGTTCGCCGCCGCTWGSGAANAATCAGAGAGCTTTATATCCAAC	1416
Db	181 AAGCGGCTATTGATGTGTGTGCAACCCCTTGGGGAATAATCGAGAGTCTGTGTCAGC	240
OY	1417 AAGGTTGCTCGCT	1429
Db	241 AGTTCATGATCCT	253
	RESULT 10	
ID	ABL76356	
XX	ABL76356 standard; cDNA: 246 BP.	
XX	ABL76356;	
XX		

DT	14-MAY-2002	(first entry)
XX	Corn tassels-derived polynucleotide (cdps) SEQ ID NO:5730.	
DE		
XX		
XX	Corn: corn tassels-derived polynucleotide; cdps; hybrid breeding; CDPs;	
KW	inherlance; characteristic; growth; development; disease resistance;	
KM	environmental adaptability; quality; yield; molecular marker;	
KX	multigene trait; plant breeding; corn tassels; gene; ss.	
XX		
OS	Zea mays.	
PN	US2001051335-A1.	
XX		
PD	13-DEC-2001.	
XX		
PF	16-APR-1999; 99US-0294093.	
XX		
PR	21-APR-1998; 98US-082567P.	
XX		
PA	(LALG/) LALGUDI R V.	
XX	(ITOL/) ITO L Y.	
PA	(SHER/) SHERMAN B K.	
PI	Lalgudi RV, Ito LY, Sherman BK;	
DR	WPI; 2002-163647/21.	
XX		
PT	Novel purified corn tassels-derived polynucleotide useful for	
PT	determining altered gene expression, to recover regulatory elements and	
PT	to follow inheritance of desirable characteristics through hybrid	
PS	breeding programs -	
XX		
PS	Claim 1; SEQ ID 5730; 201pp; English.	
XX		
CC	The present sequence describes a purified corn tassels-derived	
CC	polynucleotide sequence (cdps) comprising a nucleic acid sequence	
CC	selected from those given in ABL/0627 to ABL/7633. The cdps sequences	
CC	encode corn tassels-derived polypeptides (CDPs). The cdps sequences (I)	
CC	can be used for determining altered gene expression, to recover	
CC	regulatory elements and to follow inheritance of desirable	
CC	characteristics through hybrid breeding programs. (I) are also useful	
CC	in the evaluation, and alteration of desired characteristics associated	
CC	with growth and development, disease resistance, environmental	
CC	adaptability, quality and yield, and as molecular markers for studying	
CC	inheritance of multigene traits in a plant breeding program. (I) can be	
CC	used to produce a tassels-specific profile of gene transcription, a	
CC	transcript image, to clone regulatory elements for use in transformation	
CC	vectors, to express a polypeptide, to identify, isolate or extend	
CC	identical or related corn tassels nucleic acid sequences from DNA	
CC	libraries, in nucleic acid hybridisation or amplification technologies,	
CC	as query sequences to determine homology of known sequences, as probe	
CC	for use in Southern or Northern hybridisation, and to identify the	
CC	presence of and/or to determine the degree of similarity between two	
CC	(or more) nucleic acid sequences.	
XX		
SO	Sequence 246 BP; 58 A; 49 C; 63 G; 76 T; 0 other;	
Query Match	5.2%; Score 85.6; DB 24; Length 246;	
Best Local Similarity	60.2%; Pred. No. 1.6e-13;	
Matches 142; Conservative	0; Mismatches 94; Indels 0; Gaps 0;	
DQ	1093 TCATTCGTAAATTGGAGTGAAGCTACGCGCCCTAGTCACCTCCTGCATGCTTCCTCAA 1152	
DB	7 TCCCTAGAAGTTTGTTGGTGGGAGTGCAATCAATCATCGCTCGAGTAGACTCTCTA 66	
DQ	1153 CAGGTATCGTTTCTTCGCGCTTGTATCTTTAGAAAACCAACACTCTCGGCGTGCGTA 1212	
DB	67 CAGGAGTTGGGTTCTTCCATCGATGTTTTGTTGGTGAAGCATTAAGCTCCTTATGAGACA 126	
DQ	1213 TGTTCGGTACAACCTCCTCTCATGCTAGCTGTGCTGGAGCTATPACGAATATTTCTT 1272	
DB	127 AGTTTGGGATGACACTTGTGCTTGGGCGAGTCTATGTTGGAGCAATGCAAGAACATTTTCA 186	

CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC	and parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pcr_sequences.
XX	
SO	Sequence 10057 BP, 3023 A; 1843 C; 1967 G; 3224 T; 0 other:
Query Match	2.6%; Score 41.8; DB 22; Length 10057;
Best Local Similarity	46.7%; Pred. No. 0.99;
Matches 133: Conservative	0; Mismatches 152; Indels 0; Gaps 0;
QY	648 GGTCTGCAATGTTTGGGCTTCAAGTTGAGAGCTTCCGTTTGGAAAGGTGATGCTT 707
DB	1311 GGACGGGTAAGTCTTGTACCTCCCAATATTCACGTTGGGGGTGGAGGAATCTCT 1370
QY	708 TGGCGAATTTCTTTACGCTTTTGGATGGGTATGACTATTGTATCTGCACTGTTCTTATG 767
DB	1371 CCGTCATTTATGTTTATGATCTTTTAAAGAACAGTCAGTGAATTTTCTTTTAAAGATATAT 1430
QY	768 GCCAGTTACTGTGATCATATAGAACGATATTGACCGATTCCTGCTTATATATCCAGAA 827
DB	1431 ACCAGTGTGTCAACAACATTAATAACCATATGATCCATATATCTCTCTCA 1490
QY	828 GAATATGCAAAAGGGGAAAAAGGTCTTAACCTAAATGAATATGAAGAATAGCTTCTC 887
DB	1491 GAAATTTTAAAGTTTGTGAACCTGAAGTGTATGCAATAGATTTTATGATGCCCTTCT 1550
QY	888 TATCTTGATAGATTCCTTATATCTTATCTTTATTAATCACTCTGCTG 932
DB	1551 TATATTTGAATAATCTTTTATATGCCATATTTATTTATTTTATTT 1595
RESULT 14	
ABL70490	
ID	ABL70490 standard; DNA; 12669 BP.
XX	
AC	ABL70490;
XX	
DT	01-JUL-2002 (first entry)
XX	
DE	Chemically treated cell signalling DNA sequence complementary to#190.
XX	
KW	Cell signalling; cytosine methylation; cell signalling disease;
XX	cancer; tumour; cytosolic; ds.
OS	Unidentified.
XX	
PN	WO200202807-A2.
XX	
PD	10-JAN-2002.
XX	
PE	29-JUN-2001; 2001WO-EP07471.
XX	
PR	30-JUN-2000; 2000DE-1032529.
XX	
PR	01-SEP-2000; 2000DE-1043826.
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI: 2002-154758/20.
XX	
PT	Nucleic acid, useful for diagnosis and therapy of diseases associated
XX	with cell signalling e.g. cancer, comprises chemically modified genomic
XX	sequences of genes associated with cell signalling -
XX	
XX	Claim 1; SEQ ID NO 380; 24p+sequence listing; English.

CC	The invention relates to a nucleic acid comprising a sequence of at least
CC	18 bases of a segment of chemically pretreated DNA of genes associated
CC	with cell signalling. The activity of the modified sequences of the
CC	invention may be described as cyostatic. The object of the invention is
CC	to provide the chemically modified DNA of genes associated with cell
CC	signalling, as well as oligonucleotides and/or PN-oligoniers for
CC	detecting cytosine methylations, as well as a method which is
CC	particularly suitable for the diagnosis and/or therapy of genetic and
CC	epigenetic parameters of genes associated with cell signalling. The
CC	chemically modified DNA provided by the invention is useful for diagnosis
CC	and therapy of diseases such as solid tumours and cancer. The sequences
CC	given in records ABL70111-ABL70626 represent chemically pre-treated
CC	genomic DNA's of genes associated with cell signalling.
CC	Note: The sequence data for this patent is not represented in the printed
CC	specification, but is based on sequence information supplied by the
CC	European Patent Office.
XX	
SQ	Sequence 12669 BP; 3610 A; 224 C; 2687 G; 6148 T; 0 other;
XX	
Query Match	2.5%; Score 41.4; DB 24; Length 12669;
Best Local Similarity	49.3%; Pred. No. 1.4;
Matches 108; Conservative	0; Mismatches 111; Indels 0; Gaps 0;
OY	629 TATTTCCTTACTAGCTTCGTGCGTGCAATTTGGTGGCCTCAAGCTGACGCTTCGT 688
Db	10047 TTTTTTATTATGTTTGTGAAGCTGGTAAGTGTATATATAAAGAATGAAGTAGTAT 10106
OY	689 TTTCGAAGGTGTAGATCCCTGGGGAATTTCTTACGTTTGTGAAGGCATGACTATTGT 748
Db	10107 TTTGTGAAGGAGTGTTGATTTTAAATATTTTATTATTGTTTGTATTATTTTAA 10166
OY	749 ATCTGGACTTGTCTTATGCCCAGTCTACTGGTGATCAATAAGAACGTAATGACCGATCC 808
Db	10167 ATTGAATAATGAATGTTAATGGCTTTTAAAGAGATGTGATAGATTGATGAGGGGTGA 10226
OY	809 TCCGCTCATATCCAGAAAGAAATGCCAAAAGGCGAAAAA 847
Db	10227 TGTTTATATTAATAAAGAGTAATGAAAGTAAGCAAATA 10265
RESULT 15	
AAS61449	
ID	AAS61449 standard; DNA; 12669 BP.
XX	
AC	AAS61449;
XX	
DT	29-JAN-2002 (first entry)
DE	
Human gene regulation-associated gene oligonucleotide #404.	
XX	
Human: Gene regulation-associated gene; severe combined immunodeficiency;	
KW	cardiac damage; Inflammatory response; Haemophilia; Werner syndrome;
KW	asthma; HDR syndrome; congenital heart defect; Saelthe-Chotzen syndrome;
KW	renal disease; precleampsis; cardiac allograft vascular disease;
KW	colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW	immunostimulant; cardiant; antiinflammatory; coagulant; antislumatic;
KW	nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
OS	Homo sapiens.
XX	
WO200177375-A2.	
XX	
PD	18-Oct-2001.
XX	
PF	06-APR-2001; 2001WO-EP03968.
XX	
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
EA	(EPIC-) EPIDENOMICS AG.
XX	

PI	Olek A, Piepenbrock C, Berlin K;
XX	WPI; 2002-017470/02.
DR	New nucleic acid sequences from chemically modified genes associated
PT	with gene regulation, useful for analysing cytosine methylations for
PT	diagnosis and therapy of diseases e.g. severe combined immunodeficiency
XX	disease
PS	Disclosure; SEQ ID NO 412; 26pp; English.
XX	
CC	The invention relates to 224 nucleic acid sequences comprising at least
CC	18 bases of a chemically pretreated gene associated with gene regulation
CC	selected from 43 known genes (or complementary sequences). The
CC	chemical pretreatment converts cytosine bases unmethylated at the
CC	5-position to uracil or another base with hybridisation behaviour
CC	dissimilar to cytosine, to enable analysis of cytosine methylations.
CC	The DNA sequences, oligomers (or sets/arrays) and method are
CC	useful in the diagnosis of diseases (or predisposition to diseases)
CC	associated with gene regulation and in therapy of such diseases, by
CC	enabling analysis of the cytosine methylation patterns of such genes,
CC	kits are provided. They are especially useful in diagnosis
CC	and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC	disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC	asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC	pre-eclampsia, graft versus-host disease. The present sequence is a
CC	sequence included in the sequence data for this specification and is
CC	associated with the human gene regulation-associated genes.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pcl-sequences
XX	
CC	Sequence 12669 BP; 3610 A; 224 C; 2687 G; 6148 T; 0 other;
XX	
XX	Query Match
XX	Best Local Similarity 2.5%; Score 41.4; DB 24; Length 12669;
XX	Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY	629 TATTCTTCTACTAGCTCTGCGTCGCAATTGTTGGCTTCAAGTGGAGAGCTTCGT 688
DB	10047 TTTTCTTATATCTGTTGATTAAGCTTGGAATAGTTTATATTAATAAGTAAATAGTTAT 10106
QY	689 TTTCGAAGGTGTAGATCTCTGGGAATTTCTTACGCTTTGATGGCTATGACTATTGT 748
DB	10107 TTTGTAAGAGGAGCGTGTATATTTTTTAAATATTTTATTTTATGCTTTTATTTTGA 10166
QY	749 ATCGGAGCTGTCTTATTTGGCCACTGCTCGTGATCAATTAAGAATATGACCGATCC 808
DB	10167 ATTGAATTTAGATCTTATATGCGTTTAAAGAGATATGATTAAGCTTTGATCGAGGGTGA 10226
QY	809 TCGCTTCTATATATCCAGAGAAATGCAGAAAGGGAATA 847
DB	10227 TGTTTTATATATATAAAGAGTAATAGAAAGTAATA 10265
XX	
XX	RESULT 16
XX	ABK31523
XX	ID ABK31523 standard; DNA; 12669 BP.
XX	ABK31523;
XX	23-APR-2002 (first entry)
XX	Signal transduction associated gene modified complementary DNA #183.
XX	Human; signal transduction associated gene; cytosine methylation state;
XX	CG island; signal transduction associated disease; solid tumour; cancer;
XX	antitumour; cytostatic; mutant; ds.
XX	Homo sapiens.
XX	synthetic.

PM WO200200926-A2.
XX
XX 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07472.
XX
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
XX
XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction
XX
PS Claim 1; SEQ ID No 366; 24pp; English.
XX
XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, heart, prostate, lung, breast or liver.
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. AKK3158-ABK3155 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 12669 BP; 3610 A; 224 C; 2687 G; 6148 T; 0 other;
Query Match 2.5%; Score 41.4; DB 24; Length 12669;
Best Local Similarity 49.3%; Pred No. 1.4; Indels 0; Gaps 0;
Matches 108; Conservative 0; Mismatches 111;
OY 629 TATTTCTTACTAGCTTCTGTCGTCACATTTGTTGGCTCAAGTTGAGAGCTTCCT 688
DB 10047 TTTTATATATTTGTTGATAGAGTTGGTGTAGTTTATTAAGATGACATAGTAT 10106
OY 689 TTCTAAGAGTGTAGATCCCTGGGGAATTCCTTACGCTTTTGAGGCTATGACTAT 748
DB 10107 TTTTGAAGAGGCTGTGATTTTAAATTTTATTTTATTTTGTCTTTTATTTT 10166
OY 749 ATCTGAGACTGTTCTTATGAGCCAGTTACGTGATCATTAAGACGTATACAGCAT 808
DB 10167 ATTGAATATTAGATTGATTCGCTTTTACAGATATGATTAAGATTTTGAAGAG 10226
OY 809 TCGCTTCTATATCCAGAAAGATGCAAAAGGGAATAA 847
DB 10227 TGTTTTATATATTAAGAGTAATGAAGTAAGAAATA 10265
RESULT 17
AAK99539 standard; DNA: 1887 BP.
XX
XX AAK99539;
XX
DT 05-OCT-1999 (first entry)
XX

DE Nucleic acid sequence from U. urealyticum.
XX
XX Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;
KW human urogenital tract; pregnancy; neonatal disease; drug therapy;
KW suppurative arthritis; ss.
XX
XX Ureaplasma urealyticum.
OS WO9939007-A1.
XX
XX 05-AUG-1999.
XX
XX 29-JAN-1999; 99WO-US01972.
XX
XX 30-JAN-1998; 98US-0073189.
XX
XX (UABR-) UAB RES FOUND.
XX
XX Cassell GH, Chen EY, Glass JL, Glass JS, Heiner CR;
PI Lefkowitz E;
XX
XX WPI; 1999-469343/39.
XX
XX
XX Detection of Ureaplasma urealyticum using novel genes, probes and
PT primers
XX
PS Claim 1; Page 31-32; 110pp; English.
XX
XX The present invention provides methods for the detection and diagnosis
CC of Ureaplasma urealyticum infection. It provides novel genes
CC (AAK99501-681) that can be used as a source of primers and probes for the
CC detection and/or quantification of U. urealyticum in a biological
CC sample. The probes that can be used in the method of the invention by
CC forming target:probe complex is complementary to a region selected from
CC one of the 181 nucleic acid sequences (AAK99501-681). U. urealyticum is
CC an opportunistic pathogen of the human urogenital tract that is a
CC significant cause of adverse pregnancy outcome, neonatal disease, and
CC suppurative arthritis. As the infections are commonly asymptomatic, it is
CC important to have specific and sensitive methods for detecting their
CC presence in a patient. Also, as the pathogen has no current antibiotic
CC directed specifically against it, it would be advantageous to isolate and
CC detect gene sequences which are unique to it, and utilise these as a
CC basis for diagnosis of U. urealyticum infection as well as to develop new
CC and improved drug therapies. The present invention provides such novel
CC polynucleotide sequences (AAK99501-681).
XX
SQ Sequence 1887 BP; 600 A; 188 C; 186 G; 913 T; 0 other;
Query Match 2.5%; Score 41; DB 20; Length 1887;
Best Local Similarity 51.2%; Pred No. 0.78;
Matches 149; Conservative 0; Mismatches 135; Indels 7; Gaps 2;
OY 715 TTTCTTACGCTTTTGTGATGCTATGACATATGTATCTGACTTCTTATGCGCAGTT 774
DB 824 TCTTTTGAATATTTTGTGTTTTCATATAGATAGATGATGATGATGCTGCTTT 883
OY 775 ACTGCGATCATTAAGACGATATGACGATCTGCTTC---TATATCAGAGAA 830
DB 884 AATGATTTGAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 943
OY 831 ATGCAAAAGGAAAAAGAGCTTAACCTAAATGAATATGAAGATATGCTTCCCTAT 890
DB 944 TTTAAACCTTTGAATATATTTTGTAAATATTAATTAATTAATTAATTAATTAAT 1003
OY 891 CTTGATAGATCTCTTATATCTTTTATTAATCTCTTGGTATTCCTATGATA--TT 947
DB 1004 TCTTCAATATCTAATGATTTTAAACATTAACATTAATTAATTAATTAATTAAT 1063
OY 948 TGCATTAATCTTAATGAAGTACTTGAAAGTCAAGTCAAGTCAATAT 998
DB 1064 TGTTTTAAATTTAGTAAATTTTAAATGATCAAAATCAATTAATTAATTT 1114

ABO69222/c	
ID ABO69222 standard; DNA; 3221 BP.	
AC ABO69222;	
DT 29-AUG-2002 (first entry)	
DE Listeria monocytogenes 4b contig DNA sequence #1988.	
KW Antibacterial; Listeria; food contamination; mutational analysis;	
XX infection; ds.	
OS Listeria monocytogenes 4b.	
PX WO200228891-A2.	
PN 11-APR-2002.	
PD 04-OCT-2001; 2001WO-FR03061.	
PF 04-OCT-2000; 2000FR-0012697.	
PR (INSP) INST PASTEUR.	
PA (CNRS) CNRS CENT NAT RECH SCI.	
PI Kunst F, Glaser P;	
DR WPI; 2002-332479/37.	
XX New genomic sequences from Listeria species, useful for detection,	
PT treatment and prevention of infection, also related polypeptides,	
PT antibodies and modulators -	
PS Claim 14; SEQ ID 2035; 180bp; French.	
XX The present invention relates to nucleic acid sequences	
CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes	
CC and primers for identification and/or detection of Listeria (e.g. as	
CC contaminants in foods, or mutational analysis) and for analysis of	
CC gene expression. Proteins encoded by the nucleic acid sequences can be	
CC used to screen for compounds that modulate gene expression, replication	
CC and pathogenicity of Listeria (potential therapeutic agents), also for	
CC treating infections by Listeria, and are useful as immunogens in	
CC anti-Listeria vaccines.	
CC Note: The sequence data for this patent did not form part	
CC of the printed specification, but was obtained in electronic format	
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
SX Seq Sequence 3221 BP; 972 A; 659 C; 635 G; 955 T; 0 other:	
Query Match 2.5%; Score 40.8; DB 24; Length 3221;	
Best Local Similarity 48.7%; Pred. No. 1.1; Mismatches 117; Indels 0; Gaps 0;	
Matches 111; Conservative 0; Mismatches 117; Indels 0; Gaps 0;	
OY 224 TGGTCAGAGCAATACCTTTCATCAAGTTTGCTGTGTCGCCCTGTGCATTATCTT 283	
DB 2827 TCGTAGCAGCGTTTTCGAATTAATTAACAACGCGCATGAAGAAGCAATGCTATTATTGA 2768	
OY 284 TATGCTATTATGCAAAAGCTAAGTAATATTTTAGTAGCAGGCGCTTATTTATGCACT 343	
DB 2767 TGATGATATTATTTGAATTAACATCTCAAAATTTTTAAATAATCCAGTATTGGATTGATCCGA 2708	
OY 344 GGGAACGCCCTTTAAATTTCTTGCCCTGTGCCGACTGTAAATTTATCCGCTACGCGA 403	
DB 2707 AATTATTGCAATTTGGTGGCGATTAGCGCTCAACCAGTGTGTTGATAGCTAAACGA 2648	
OY 404 TGTTTTACATCCACAGATTGCTGACCGTTTACAGGCAACCTCAC 451	
DB 2647 AGCTGTGCGAATAAAGATGCTAATGCTTATCATGCGGCAACAAC 2600	

AB071071/C
ID AB071071 standard; DNA; 6921 BP.
XX AC ABO71071;
XX DT 29-AUG-2002 (first entry)
DE Listeria monocytogenes 4b contig DNA sequence #1013.

KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX

Listeria monocytogenes 4b.
XX WO200228891-A2.
PN 11-APR-2002.
PD

PF 04-OCT-2001; 2001WO-FR03061.
PR 04-OCT-2000; 2000FR-0012697.
XX PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;

Pt PI Kunst F, Glaser P;
DR WPI: 2002-332479/37.

New genomic sequences from Listeria species, useful for detection,
Pt treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
PS Claim 14; SEQ ID 3884; 180bp; French.

The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (*e.g.*, as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
used to screen for compounds that modulate gene expression, replication
and pathogenicity of Listeria (potential therapeutic agents), also for
treating infections by Listeria, and are useful as immunogens in
anti-Listeria vaccines.

Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 6921 BP; 1997 A; 1409 C; 1324 G; 2191 T; 0 other:

RESULT_20
ABL33000

ID	ABJ33000	standard; DNA: 6557 BP.
XX	AC	
XX	ABL33000:	
XX	DT	
XX	26-MAR-2002	(first entry)
DE	Human immune system associated gene	SEQ ID NO: 973.
XX	Human: immune system disease; cytosine methylation; antiasthmatic; antiarteriosclerotic; antianemic; cytosolic; noctropic; antiprotective; anti-HIV; anticonvulsant; ophthalmological; antirheumatic; antiarthritic; antidiabetic; antiparasitic; antinflammatory; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.	
KW	Homo sapiens.	
XX	OS	
XX	MO200200928-A2.	
XX	PN	
XX	03-JAN-2002.	
XX	PD	
XX	02-JUL-2001; 2001WO-EP07537.	
XX	PF	
XX	30-JUN-2000; 2000DE-1032529.	
XX	PR	
XX	01-SEP-2000; 2000DE-1043826.	
XX	PA	(EPIG-) EPIGENOMICS AG.
XX	PI	Olek A, Piepenbrock C, Berlin K;
XX	PI	WPI: 2002-130909/17.
DR	XX	
XX	PT	Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation
XX	PS	
XX	XX	Claim 1; SEQ ID NO 973; 32pp + Sequence Listing; German.
CC	XX	The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.
CC	XX	
XX	XX	Sequence 6557 BP: 1528 A; 211 C; 1744 G; 3073 T; 1 other:
XX	XX	
XX	XX	Query Match 2.4%; Score 39.4; DB 24; Length 6557;
XX	XX	Best Local Similarity 48.4%; Pred. NO. 3.7;
XX	XX	Matches 109; Conservative 0; Mismatches 116; Indels 0; Gaps 0
XX	XX	
XX	XX	905 TTATATCTTTTATTAACCTCTTGTTATTCCTATGATGGTATTTGCATTAACTTAATGCA 964
XX	XX	DB 3649 TTCGATTTATTTAAAGAAATTTTTTTTTTTTAAATTAATTAATTAATTTTAAATA 3708
XX	XX	965 AGTCATTTGGAAGATCGTGAACCTGCAATATCCTAATATGATGACTATATGAGTT 1024
XX	XX	DB 3709 TTTTATTTTAAAGTTTTCATTAATATTTATTTATTTATTTATTTATTTAAACT 3768
XX	XX	97 1025 CATGGGAACTTCCTCTCGACTGCGCTAGTATTCCTACTTAATCAGTATTTGGTG 1084
XX	XX	DB 3769 TTTTTCATATAATTTTTTAACTTTGGCGCTTTAAGATATGATTTGTTTTTGGAG 3828
XX	XX	97 1085 TGGTAAAGCATTTGTAATTTGAGTGAATTAAGTAACTGAGCCCAAGT 1129
XX	XX	DB 3829 TATTTAGGTATTTTACAGTTAGGTGTTATTTGGTTAATTAAGT 3873

AAS45469/C	ID	AAS45469 standard; DNA; 8588 BP.
XX	AC	AAS45469;
XX	DT	18-DEC-2001 (first entry)
XX	DE	Chemically pretreated complementary DNA associated with cell cycle #87.
XX	KX	Cell cycle: human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW	KW	human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KM	KM	graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KM	KM	arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KM	KM	immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
XX	PCR	primer.
OS	OS	Homo sapiens.
XX	PN	WO200168911-AZ.
XX	PD	20-SEP-2001.
XX	PF	15-MAR-2001; 2001WO-EP02945.
XX	PR	15-MAR-2000; 2000DE-1013847.
PR	PR	06-APR-2000; 2000DE-1019058.
PR	PR	07-APR-2000; 2000DE-1019173.
PR	PR	30-JUN-2000; 2000DE-1032529.
PR	PR	01-SEP-2000; 2000DE-1043826.
PA	PA	(EPiG-) EPIGENOMICS AG.
XX	PI	Olek A. Piepenbrock C, Berlin K:
XX	XX	WPI: 2001-602751/68.
XX	PT	Designing primers and probes for analysing diseases associated with
PT	PT	cytosine methylation state e.g. arthritis, cancer, aging,
PT	PT	arteriosclerosis comprising fragments of chemically modified genes
PT	PT	associated with cell cycle -
PS	PS	Claim 1: SEQ ID No 174; 28pp; English.
XX	XX	
CC	CC	Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC	CC	molecules associated with the cell cycle and specific PCR primers of the
CC	CC	invention. The sequences are useful for detecting the methylation state
CC	CC	of all CpG dinucleotides in a sequence and therefore for analysing
CC	CC	associated diseases. By analysing cytosine methylations in the pretreated
CC	CC	DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC	CC	of existing diseases or the predisposition to specific diseases can be
CC	CC	ascertained. The parameters may be compared to another set of genetic
CC	CC	and/or epigenetic parameters, the differences serving as basis for
CC	CC	diagnosis and/or prognosis events which are disadvantageous to patients.
CC	CC	The sequences of the invention are useful for the diagnosis and therapy
CC	CC	of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC	CC	aging, glomerular disease, Lewy body disease, arthritis,
CC	CC	arteriosclerosis, solid tumours and cancers.
XX	XX	
SQ	SQ	Sequence 8588 BP; 2142 A; 245 C; 2070 G; 4122 T; 9 other:
Query Match	2.4%;	Score 39.2; DB 22; Length 8588;
Best Local Similarity	66.7%;	Pred. No. 4.8;
Matches	56; Conservative	0; Mismatches 28; Indels 0; Gaps 0.
QY	2	AAATATAAACAATCATAGAGTAAGAAATTAAGTATTTCAGACGTAAATATGCACAANAAC 61
Db	3945	AAAAAAAAAAAAATTAATAAAAAAAAAAATAAAGCAAAAACAAAAATTAATTAATAAAAAAAAAA 3886
QY	62	CGAACAAAAACCTTTGGAAATT 85
DB	3885	CGAAAAAAACC CGCGTAAAAAATT 3862

RESULT 22
ABK28325/c
ID ABRK28325 standard; DNA: 8588 BP.
XX
AC ABRK28325;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated genomic DNA #100.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP03973.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI: 2002-090046/12.
XX
XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer -
XX
PS Claim 1: SEQ ID No 199; 32pp; English.
XX
XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABRK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 8588 BP; 2142 A; 245 C; 2070 G; 4122 T; 9 other;
Query Match 2.4%; Score 39.2; DB 24; Length 8588;

Best Local Similarity 66.7%; Pred. No. 4.8;
Matches 56; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
OY 2 AAATTAATACTATCAGATGAAATTAAGTATTTAGAGGGTAATATGACAAAAC 61
||| ||||| ||| ||||| ||||| ||| ||||| ||||| |||||
Db 3945 AAAAAAAAAATAAAAAAAAAAAAAAAAAACAAAACAAAATAATTAATAAAAAA 3886
OY 62 CGAAGAAAAACCTTTTGGAAATT 85
||| ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 3885 CGAAAAAAACCGGTAAAAAATT 3862
RESULT 23
AAD22126
ID AAD22126 standard; DNA: 1231 BP.
XX
AC AAD22126;
XX
DT 12-FEB-2002 (first entry)
XX
DE Lycopersicon esculentum sucrose transporter 1 (SUT1) gene 3' UTR.
XX
KW Tomato; sucrose transporter 1; SUT1; transgenic plant; plant growth;
KW UTR; untranslated region; gene expression; regulatory element; ds.
XX
OS Lycopersicon esculentum.
XX
PN EP114915-A1.
XX
PD 31-OCT-2001.
XX
PF 28-APR-2000; 2000EP-0109218.
XX
PR 28-APR-2000; 2000EP-0109218.
XX
XX (FROM/) FROMMER W B.
XX
PA Fischer W, Frommer WB, Hirner B, Lalonde S, Okumoto S, Tegeder M;
XX
PI Ward J, Weise A;
XX
DR WPI: 2002-012576/02.
XX
XX Regulatory element for modifying gene expression in transgenic plants
PT e.g. enhance plant growth, comprises a promoter and a nucleotide
PT enhancer or repressor having a non-coding region of a sucrose
PT transporter gene -
XX
XX
PS Claim 9; Page 17; 35pp; English.
XX
XX The invention relates to a regulatory element for the modification of
CC gene expression in transgenic plants. The regulatory element comprises a
CC promoter and a nucleotide sequence which is a nucleotide enhancer and/or
CC repressor sequence having a non-coding region of a sucrose transporter
CC (SUT) gene. The nucleotide enhancer sequence of is useful for enhancing
CC a promoter and nucleotide repressor sequence is useful for inhibiting a
CC promoter. The regulatory element is useful to enhance or inhibit
CC expression of a plant gene which includes enhancing carbon dioxide uptake
CC in plant, enhance sucrose phloem loading capacity in plants, increase
CC sugar content or oil content in plant tissues or organs or increase
CC amino acid and/or protein content in plant tissues or organs. The
CC present sequence is Lycopersicon esculentum sucrose transporter 1
CC (SUT1) gene 3' untranslated region (UTR).
XX
XX Sequence 1231 BP; 403 A; 156 C; 213 G; 459 T; 0 other;
Query Match 2.4%; Score 39; DB 24; Length 1231;
Best Local Similarity 47.3%; Pred. No. 2.3;
Matches 150; Conservative 0; Mismatches 165; Indels 2; Gaps 1;
OY 616 GTATCGGAGCTAATATTCTTACTAGCTTCGTGCGCAATGTTGGGCTCAAGT 675
||| ||| ||||| ||| ||| ||||| ||||| ||||| ||||| |||||
Db 184 GTTTCCTTTTAATTCTCCAAACAATGGGATGTGTGTTGTTCAATTTGTGTG 243

QY 676 TGAGACCTCCGTTTCGAGGTGAGATCCTTGGGGAATTTCTTACGCTTTTGATNG 735
 Db 244 TACCAAGAAAGATGTCAGTATTAACATGCTGGCAATGCTTTGACCTTTACTA 303
 QY 736 CATACGATTTGATCTGGACTGTTCTTATATGCCAGTACTGCGGATCATTAAGACG 795
 Db 304 TTTTATTTTATTTTGGTAATTTTCTTGGAGGTAAGAAAAAAGAAAGTGGT 363
 QY 796 TATTACGACCTCGCTCTTAAATCCAGAAATGCAAAAGGGAAGAAAGGTGCTA 855
 Db 364 TGGAGGTGAAGATGGGTACTTTAATTGG--CTTGCTAAGAAAGGAGATTTATGAAGA 421
 QY 856 AACCTAAATGAATGAAGATGACCTTCCTATCTGATGATCTCCTTATATCTTT 915
 Db 422 ACTTTGACTTAAGTGAAGTCAAAATGATCTATCTATATATCAATGTTTCTAT 481
 QY 916 TATTACCTCTCTGGTT 932
 Db 482 TTTTGTGTTCTTCTTT 498

RESULT 24

AAV82021/C
 ID AAV82021 standard; DNA: 2718 BP.

AAV82021;

21-JUN-1999 (first entry)

DE Moraxella catarrhalis Vh19 lbpB gene.

XX Lactoferrin receptor; lactoferrin binding protein; LBP1; LBP2;
 KW lbpA gene; lbpB gene; ORF3; infection; otitis media; sinusitis;
 KW conjunctivitis; pneumonia; bronchitis; tracheitis; emphysema;
 KW diagnosis; therapy; vaccine; Branhamella catarrhalis; ss.

OS Moraxella catarrhalis.

XX Key Location/Qualifiers
 FH 1.2718
 FT CDS

FT //tag= a
 FT /transl_except= (pos:844..846, aa:Thr)
 FT /transl_except= (pos:847..849, aa:Thr)
 FT /transl_except= (pos:850..852, aa:Thr)
 FT /transl_except= (pos:853..855, aa:Thr)
 FT /transl_except= (pos:856..858, aa:Thr)
 FT /transl_except= (pos:859..861, aa:Thr)
 FT /transl_except= (pos:862..864, aa:Thr)
 FT /transl_except= (pos:868..870, aa:Thr)
 FT /transl_except= (pos:871..873, aa:Thr)
 FT /transl_except= (pos:874..876, aa:Thr)
 FT /transl_except= (pos:877..879, aa:Thr)
 FT /transl_except= (pos:880..882, aa:Thr)
 FT /transl_except= (pos:886..888, aa:Thr)
 FT /transl_except= (pos:891..891, aa:Thr)
 FT /transl_except= (pos:892..894, aa:Thr)
 FT /transl_except= (pos:895..897, aa:Thr)

XX W09855606-A2.

XX 10-DEC-1998.

XX 02-JUN-1998; 98WO-CA00544.

XX 08-MAY-1998; 98US-0074658.

XX 03-JUN-1997; 97US-0867941.

XX (CONN-) CONNAUGHT LAB LTD.

XX Du R, Klein MH, Loosmore SM, Wang Q, Yang Y;

XX WPI: 1999-070266/06.

XX P-PSDB; AAM89421.

XX Lactoferrin receptor genes from Moraxella, especially M. catarrhalis
 PT - useful to diagnose Moraxella infection e.g. to detect otitis media
 PT due to M. catarrhalis infection and to immunise against such
 PT infections

PS Claim 8; Fig 16A-O; 202pp; English.

XX This polynucleotide comprises the lactoferrin binding protein lbpB
 CC gene of Moraxella catarrhalis (Branhamella catarrhalis) strain Vh19.
 CC It encodes lactoferrin binding protein 2 (lbp2, see AAM89421). The
 CC genes and DNA sequences of the lactoferrin receptor (lfr) locus of
 CC Moraxella are useful for diagnosis, immunisation, and the
 CC generation of diagnostic and immunological reagents. Immunogenic
 CC compositions, including vaccines, based upon expressed recombinant
 CC lbp1 and/or lbp2 and/or ORF3, portions of these, or their
 CC analogues, can be prepared for prevention of diseases caused by
 CC Moraxella. M. catarrhalis is a causative agent of otitis media and
 CC has been associated with sinusitis, conjunctivitis and inflammatory
 CC diseases of the lower respiratory tract, such as pneumonia, chronic
 CC bronchitis, tracheitis and emphysema.

XX Sequence 2718 BP; 958 A; 585 C; 547 G; 628 T; 0 other;

Query Match 2.4%; Score 38.8; DB 20; Length 2718;

Best Local Similarity 57.4%; Pred. No. 3.7; Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1079 TGTGTGTGTAACGATTCGTAATTTGGATGTTAACTGAGCCCTAGTCACTCTGT 1138

Db 1231 TATTGCTGTTATCATCAATCAAAAGATTGGCTGCGAGTAAGGGTGGCATTAACT 1172

QY 1139 CATGTTCTCCCAACAGGATGTTTTCCTCTGTTATCTTTAGAACCAAGCTTC 1198

Db 1171 TGGCTCAATCTCATGACATCGCTTCATTGACCTTTTGGGATCAACAGTCTTC 1112

QY 1199 TG 1200

Db 1111 TG 1110

RESULT 25

AAH48024/C
 ID AAH48024 standard; DNA: 14041 BP.

AAH48024;

DT 18-SEP-2001 (first entry)

XX Internal control B19c #1.

KW Internal control; ss.

XX Parvovirus.

XX W0200146463-A2.

XX 28-JUN-2001.

XX 20-DEC-2000; 2000WO-EPI2996.

XX 22-DEC-1999; 99AT-0002170.

XX (BAXT) BAXTER AG.

XX Zimmermann K, Turecek P, Schwarz H, Rieger M;

XX WPI: 2001-408658/43.

XX Internal standards useful for nucleic acid amplification assays,
 PT comprises a synthetic nucleic acid made by non-recombinant techniques

PI Tectelin H;
XX
XX WPI: 2002-352536/38.
DR P-PSDB: ABP28137.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 7; Page 3706; 4525bp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and anti-inflammatory
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 924 BP; 280 A; 159 C; 211 G; 274 T; 0 other;
XX
Query Match 2.4%; Score 38.6; DB 24; Length 924;
Best Local Similarity 49.3%; Pred. No. 2.6;
Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 558 GGAATTGCTAATGAATACAAATCAAGAACGCTTCTACGCTCTTTTCGT 617
DB 181 GAATTTGATGCTAAATCAAGAGTTGATGACAGAAATCGCTTCATCTTCACGT 240
QY 618 ATCGAGCTAATATTTCTTACTAGCTTCTGCTGCAATTTGTTGGCTTCAAGTTG 677
DB 241 CGTGAAGTTTGAAGCAATCAGCAGCTGCACGCTTAAAGATTTTCAAAACATTTGAA 300
QY 678 AGAGCTTCGCTTCTCAAGGTGATGATCCTTGGGCAATTTCTTACGCTTTTGGATGCGT 737
DB 301 GTAGGTTGAGTTGTAAGTGTGTAAGTTGCTCGTTTAAACAGCTTTGCTCTTCAATTGAC 360
QY 738 ATGACTATTGTATCTGAGCTGTTC 762
DB 361 CTTGGTGGTGTGATGAGCTTGTAC 385
XX
RESULT 28
AAH52921
ID AAH52921 standard; DNA; 1176 BP.
XX
AC AAH52921;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1235.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX

PR 09-NOV-1999; 99US-0164258.
XX
XX (GLAXO) GLAXO GROUP LTD.
XX
XX
PI Kimerly WJ;
XX
XX WPI: 2001-316495/33.
DR P-PSDB: AAG82071.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
PT
PS Claim 8; Page 354; 2188bp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 1176 BP; 342 A; 164 C; 219 G; 451 T; 0 other;
XX
Query Match 2.4%; Score 38.6; DB 22; Length 1176;
Best Local Similarity 49.3%; Pred. No. 2.9;
Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 1054 TAGTATCCCTACTATATCATCTATTTGTTGGGTGATGCTATTCGTAATTTGATGCT 1113
DB 827 TACTATTGTATCAATCATACATATTTGTTAGCTTATTTCAAAATTCATATTAAGCTTAA 886
QY 1114 TAACGTGAGCCCTAGTCACTCCTGTCATGCTTCTCTAAGAGTATGTTTCTTCGCTC 1173
DB 887 TCATTTGCTTCTATTCGGTGTGATGATTTGGTACACACCGTACTTAATAGCAAAA 946
QY 1174 TTGTTATCTTTAGAAACCAAGCTTCTGGGCTGCTGCTATGTTGCTGTAACACTCCCTCA 1233
DB 947 TTATCTTAAAGTCGCAAGACACCTTCTTCAAGTACGTTAGCTGCTCTATTTTCA 1006
QY 1234 TCGTAGCTGTGTTGTCGAGCTAT 1258
DB 1007 ATGTTGCTAATTTTCTTGTGTCAT 1031
XX
RESULT 29
ABN70725
ID ABN70725 standard; DNA; 1398 BP.
XX
AC ABN70725;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 9363.
XX
XX Streptococcus polynucleotide
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus agalactiae.
XX

AAH54155/c
ID AAH54155 standard; DNA: 3332 BP.
XX
AC AAH54155;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3519.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
KM vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimerly WJ;
XX
DR WPI: 2001-316495/33.
XX
PT Nucleic acid encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1094-1095; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 3332 BP: 1238 A; 494 C; 497 G; 1103 T; 0 other;
XX
Query Match 2.4%; Score 38.6; DB 22; Length 3332;
Best Local Similarity 49.3%; Pred. No. 4.6;
Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

DB 1225 ATGTTGCTAATTTCTTGTCGTCAT 1201
|||||
RESULT 32
ABL32869
ID ABL32869 standard; DNA: 5642 BP.
XX
AC ABL32869;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 842.
XX
KM Human immune system disease; cytosine methylation; antiasthmatic;
KM antiarteriosclerotic; antihaemic; cyostatic; noctropic;
KM neutroprotective; anti-HIV; anticonvulsant; ophthalmological;
KM antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KM gene; ds.
XX
XX Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 842; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/alcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5642 BP: 1272 A; 316 C; 1585 G; 2469 T; 0 other;
XX
Query Match 2.4%; Score 38.6; DB 24; Length 5642;
Best Local Similarity 47.0%; Pred. No. 5.8;
Matches 119; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Oy	727	TTTGGATGGCATACATCATCTGTATCTGAGACTTCTCTTATGCCACTACTGCTGGATCA	786
Db	1548	TTTGAGGCCGAATATTATGTTGATGTTGATTAATTTATTTGCGATTAGGAATGATG	1607
Oy	787	ATAAGACGTAAT	799
Db	1608	TGAGAGAGCGAAT	1620
RESULT 33			
ABL92259			
ID	ABL92259	standard; DNA; 7695 BP.	
AC	ABL92259;		
XX			
DT	01-JUL-2002	(first entry)	
XX			
DE	Chemically treated DNA repair gene fragment complementary to#34.		
XX			
KW	DNA repair; cytosine methylation; PMS2L1, PMS2L12, PMS2L3;		
KW	PMS2L4; PMS2L5; PMS2L6; MGMT; MSH2; NDDT1; TDG; INPPL1; RFC4;		
KW	DDT1L1; FANCB; XRCC8; ataxia telangiectasia; aging; Bloom's syndrome;		
KW	Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;		
KW	immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;		
KW	cancer; ds.		
XX			
OS	Unidentified.		
XX			
PN	WO200181622-A2.		
XX			
FD	01-NOV-2001.		
XX			
PF	06-APR-2001; 2001WO-EP03972.		
XX			
PR	06-APR-2000; 2000DE-1019058.		
PR	07-APR-2000; 2000DE-1019173.		
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPig-) EPIGENOMICS AG.		
PI	Olek A, Piepenbrock C, Berlin K;		
DR	WPI; 2002-034446/04.		
XX			
PT	New nucleic acid derived from genes associated with DNA repair, useful		
PT	for diagnosis, e.g. of ataxia telangiectasia, by determination of		
PT	cytosine methylation -		
XX			
PS	Claim 1; SEQ ID NO 68; 25bp + sequence listing; English.		
CC			
CC	The invention relates to nucleic acids containing a sequence of at least		
CC	18 nucleotides of chemically treated DNA of genes associated with DNA		
CC	repair, and their complements. The invention also relates to nucleic		
CC	acids comprising at least 18 base pairs of the chemically pretreated DNA		
CC	of genes associated with DNA repair selected from PMS2L1, PMS2L12,		
CC	PMS2L3, PMS2L4, PMS2L5, PMS2L6, MGMT, MSH2, NDDT1, TDG, INPPL1,		
CC	RFC4, DDT1L1, FANCB, or XRCC8. Nucleic acids of the invention and related		
CC	oligonucleotides, are useful for diagnosis of diseases associated with gene		
CC	repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,		
CC	Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,		
CC	immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours		
CC	and/or by detecting single-nucleotide polymorphisms. Determination of		
CC	individual methylation patterns may allow development of individualised		
CC	therapies. The sequences given in records ABL92192-ABL92335 represent		
CC	chemically pre-treated DNA fragments from genes associated with DNA		
CC	repair, and their complements.		
CC	Note: The sequence data for this patent is not represented in the		
CC	specification, but is based on sequence information supplied by the		
CC	European Patent Office.		
XX			
XX	Sequence 7695 BP; 2111 A; 268 C; 1887 G; 3429 T; 0 other;		

Query Match	2.43;	Score 38.6;	DB 24;	Length 7695;
Best Local Similarity	43.33;	Pred. No. 6.7;	Mismatches 234;	Indels 0;
Matches 179;	Conservative	0;	Mismatches	Gaps
QY 560	ATTGCTAATGAATTTACAAAATATCCAGCAAGCAAAAGCGTTCTACGCTCTTTCGGTAT	619		
Db 5162	ATTTTAAATTAATTTTATAGAAAAGAAAAGAAAATATTATTATTTTTCCTTTT	5241		
QY 620	CGGAGCTAATATTTCTTACTAGCTTCTGTCGCAATTTGTTGGCTTCAAGTTGAG	679		
Db 5242	TGATAGCATATATTATTAATTTTGTCGTTTATAGTGTGTTTAAAGTTAT	5301		
QY 680	AGCTTCGCTTCTCAAGGTAGATCCCTGGGAAATTTCTTAACGTCCTTATGCGTAT	739		
Db 5302	ATGTTTAATATAGTTTTTTTTTTTATTTAAAGACATTTTTTTAGCGTTTATGATTTTT	5361		
QY 740	GACATATTTATCTGACCTGTCTTATGGCCAGTACTGCTGATCAATAAGACGTAT	799		
Db 5362	GCTTAAAGCAAGAAATATGGTATATATTTATGATTTTTTATATATGTAACGAT	5421		
QY 800	GACGATCTCGCTTCTATATATCCAGCAAGAAATGCCAAAAGGGAAAAAGTGTCTAAAC	859		
Db 5422	TGGGAGGTAAATTAATATTTTTTAAATTTAAAGTTCAGGTTTAAATAAGTTAATAG	5481		
QY 860	TAAATGAATTTGAAGATATAGCTCCCTATCTTGATAGATCCCTATATCTTTAT	919		
Db 5482	TAAATTAATATTTGTTATTTTATTTTATTTTATTTTATGATTAATGTTATGTGAGTATTTTAA	5541		
QY 920	AACCTCTTGTTATTTGCTGATAGATTTTTCGATTAACCTTAATGCAAGTACTT	972		
Db 5542	ATGCAATTTGCAATATATTAATTAATTTTATTTGATTAATAATTTTGAATGATTAAT	5594		

RESULT 34	
ABL33537	
ID	ABL33537 standard; DNA: 7695 BP.
XX	
AC	ABL33537:
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 1510.
XX	
KW	Human; immune system disease; cytosine methylation; antiasthmatic;
KW	antiarteriosclerotic; antiataemic; cytosatic; noctropic;
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW	antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW	antiflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
gene; ds.	
XX	
OS	Homo sapiens.
XX	
PN	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
XX	
PP	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIC-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
XX	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation -

XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tetteijn H;
 XX
 DR WPI: 2002-352536/38.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS
 PS Claim 8: Page 4196-4488; 4525pp; English.
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;
 Query Match 2.4%; Score 38.6; DB 24; Length 2155561;
 Best Local Similarity 49.3%; Pred. No. 79;
 Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
 QY 558 GGATTTGCTAATGAATATACAAATCCAGCAAGCAAGCGTTTCTAGCTCTTTTCGGT 617
 Db 1124096 GAATTTATGCTAATAATCAAGAAGATGATGACACAGAAATCCCTTATCTTTCACGT 1124037
 QY 618 ATCGAGCTAATATTTCTTACTAGCTTCTGCTGCTGCAATGTTGGGCTTCAAGTTG 677
 Db 1124036 CGTGAACGTTGTGAGATGACGACGCTGCAAGCAAGTTTCTCAAAACATTGAA 1123977
 QY 678 AGAGCTTCGTTTCTGAGAGGTGTAGATCCCTGGGAATTTCTTACGCTTTGATGGCT 737
 Db 1123976 GTAGGTTCACTGTGAACCTGTAAGTTGCTGTTTAAACAAGCTTTGGTCTTCATTGAC 1123917
 QY 738 ATGACTATTGTATCTGCACTTGTTC 762
 Db 1123916 CTTGGTGTCTGTGATGACCTTGTAC 1123892
 RESULT 37
 ABN91748
 ID ABN91748 standard; DNA; 1269 BP.
 XX
 AC ABN91748;
 XX
 XX 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1211.
 XX
 KW Staphylococcus epidermidis: open reading frame: ORF; bacterial infection;
 KW antibacterial; gene therapy; gene; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN US6380370-B1.
 XX

PD 30-APR-2002.
 XX
 XX 13-AUG-1998; 9805-0134001.
 XX
 PR 14-AUG-1997; 9705-055779P.
 PR 08-NOV-1997; 9705-064964P.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Doucette-Stamm LA, Bush D;
 XX
 DR WPI: 2002-381255/41.
 DR P-PSDB: ABP39203.
 XX
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX
 PS Disclosure; SEQ ID 1211: 267pp; English.
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 XX
 SQ Sequence 1269 BP; 381 A; 176 C; 229 G; 483 T; 0 other;
 Query Match 2.3%; Score 38.4; DB 24; Length 1269;
 Best Local Similarity 47.5%; Pred. No. 3.4;
 Matches 114; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 QY 1019 TGAGTTCAVGGGAACCTTCCTCTGAGCTGGCGTAGTATCCGTAATATATGCTAAT 1078
 Db 885 TAAATTACAAATAATTAATTAACACATTTCTTTACTAATTTGTAACAATCATATAT 944
 QY 1079 TCTTGTGCTAGACGTCATCTGTAATTTGGATGGTTAACTGAGACCCCTAGTACTCTGT 1138
 Db 945 TGTTACTGTGTATCAAAATTCATAATTAATGATTAATCATTTTGCTTCTTATTCGGTTTGG 1004
 QY 1139 CATGTTTCTTCTACAGGATATCGTTTCTGCTTGTATCTTTAGAACCAAGCTTC 1198
 Db 1005 TACATTTGTGTAACACACCGTTACTTAATAGCAAAATTTCTTAAGTGCAAAAGAAGCACC 1064
 QY 1199 TGGCGTGTGCTATGTTGCTAGCAACTCTCTCATGCTAGCTGTGTTGTGGAGCTAT 1258
 Db 1065 ACTTCTTGCAGTAGTAGGAGCTGTATTTCAATGTCTTAATTTTCTTGTGTCGAAT 1124
 RESULT 38
 ABQ68092
 ID ABQ68092 standard; DNA; 903 BP.
 XX
 AC ABQ68092;
 XX
 XX 29-AUG-2002 (first entry)
 XX
 DE Listeria monocytogenes EGD DNA sequence #216.
 XX
 KW Listeria monocytogenes EGD: food contamination; mutational analysis;
 KW infection; ds.
 XX
 OS Listeria monocytogenes EGD.
 XX
 PN WO200228891-A2.
 XX
 PD 11-APR-2002.
 XX

PF 04-OCT-2001; 2001WO-FR03061.
 XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX WPI; 2002-332479/37.
 DR
 XX New genomic sequences from *Listeria* species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 XX
 PS Claim 16; SEQ ID 905; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
 CC and primers for identification and/or detection of *Listeria* (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
 CC treating infections by *Listeria*, and are useful as immunogens in
 CC anti-*Listeria* vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 903 BP; 283 A; 131 C; 226 G; 263 T; 0 other;

Query Match 2.3%; Score 38.2; DB 24; Length 903;
 Best Local Similarity 48.0%; Pred. No. 3.3; Mismatches 118; Indels 0; Gaps 0;
 Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
 OY 225 GGTGAGAGCAATACCTTCATCAAGTTTGGCTGTGTCCTGCTATATCTTT 284
 DB 583 GGTGAGAGCAATACCTTCATCAAGTTTGGCTGTGTCCTGCTATATCTTT 284
 OY 285 ATGCTATATTTATGCAAGTAAAGTAATATTTAAAGTAAGCAAGCCCTATTTATGCGATG 344
 DB 643 GGTATATTTATGAAATTTAGCATCTCAAAATTTAAATATCCAGATTTGATGATCGGAA 702
 OY 345 GGAAGCCCTTTTATTTTCTTGGCCGTGTCGCCAGCTATATTTATGCGATG 404
 DB 703 ATTATTTGCAATTTGGTGGCATTTAGCGCCCAACCGGTGTTGTAAGGCTAAAGCAT 762
 OY 405 GTTTTACATCTACAGAAATTTGCTGACGTTTACAGGCGCATCTTACC 451
 DB 763 GCTGTGCGAAGATAAAGACAGCTATCTTATCATGCCGACACACC 809

RESULT 39
 AB069810
 ID AB069810 standard; DNA; 903 BP.
 XX
 AC AB069810;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE *Listeria* monocytogenes EGDe DNA sequence #22.
 XX
 XX Antibacterial; *Listeria*; food contamination; mutational analysis;
 KW infection; ds.
 XX
 OS *Listeria* monocytogenes EGDe.
 XX
 PN WO200228891-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 04-OCT-2001; 2001WO-FR03061.

XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX WPI; 2002-332479/37.
 DR
 XX New genomic sequences from *Listeria* species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 XX
 PS Claim 16; SEQ ID 2623; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
 CC and primers for identification and/or detection of *Listeria* (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
 CC treating infections by *Listeria*, and are useful as immunogens in
 CC anti-*Listeria* vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 903 BP; 283 A; 131 C; 226 G; 263 T; 0 other;

Query Match 2.3%; Score 38.2; DB 24; Length 903;
 Best Local Similarity 48.0%; Pred. No. 3.3; Mismatches 118; Indels 0; Gaps 0;
 Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
 OY 225 GGTGAGAGCAATACCTTCATCAAGTTTGGCTGTGTCCTGCTATATCTTT 284
 DB 583 GGTGAGAGCAATACCTTCATCAAGTTTGGCTGTGTCCTGCTATATCTTT 284
 OY 285 ATGCTATATTTATGCAAGTAAAGTAATATTTAAAGTAAGCAAGCCCTATTTATGCGATG 344
 DB 643 GGTATATTTATGAAATTTAGCATCTCAAAATTTAAATATCCAGATTTGATGATCGGAA 702
 OY 345 GGAAGCCCTTTTATTTTCTTGGCCGTGTCGCCAGCTATATTTATGCGATG 404
 DB 703 ATTATTTGCAATTTGGTGGCATTTAGCGCCCAACCGGTGTTGTAAGGCTAAAGCAT 762
 OY 405 GTTTTACATCTACAGAAATTTGCTGACGTTTACAGGCGCATCTTACC 451
 DB 763 GCTGTGCGAAGATAAAGACAGCTATCTTATCATGCCGACACACC 809

RESULT 40
 ABL33669
 ID ABL33669 standard; DNA; 5347 BP.
 XX
 AC ABL33669;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1642.
 XX
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antiamebic; cytosolic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX

```

PN WO200200928-A2.
XX
XX
PD 03-JAN-2002.
XX
XX PF 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
XX
PS Claim 1; SEQ ID NO 1642; 32pp + Sequence Listing; German.
XX
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5347 BP; 1429 A; 60 C; 1153 G; 2705 T; 0 other;

Query Match 2.3%; Score 38.2; DB 24; Length 5347;
Best Local Similarity 52.8%; Pred. No. 7.3;
Matches 105; Conservative 0; Mismatches 93; Indels 1; Gaps 1;

QY 737 TATGACTATGTATCGACTGTCTTATGCGCCAGTTACTGCGATCAATPAGACGT 796
   ||| || ||| || || || || || || || || || || || || || || || ||
DB 1653 TATTGGAATGATATATAGCTTTTATTTTATGAAATTTAATTAGATTTTTGTGTCG 1712
   || || || || || || || || || || || || || || || || || || || ||
QY 797 ATTGACCGATCGCTCTATATCCAGAGAAATGCAAAAGCGAAAGTGCTAA 856
   || || || || || || || || || || || || || || || || || || || ||
DB 1713 TTTTGAGAGATTTTGTATAGTAGATTAGTAAGTAATTAGTAATATGATGTTAA 1772
   || || || || || || || || || || || || || || || || || || || ||
QY 857 ACCTAAATGATATGAAAGATAGCTTCCCTATCTTGATAGATCCTTATATCTTTT 916
   ||| || ||||| ||| || || || || || || || || || || || || || ||
DB 1773 TGAGAAATTAGTATGAATATAATTTTTTTGTGTTTATATGATATGGGA-TTTT 1831
   || || || || || || || || || || || || || || || || || || || ||
QY 917 ATTAACCTCTTGTTATT 935
   ||||| || || |||||
DB 1832 ATTAATTTTATTTATT 1850
   ||||| || || |||||

```

Search completed: January 21, 2003, 19:15:44
 Job time : 4255 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 19:33:03 : Search time 253 Seconds
(without alignments)
14571.230 Million cell updates/sec

Title: US-09-869-433-1

Perfect score: 1637

Sequence: 1 gaataaanaactatcagaa.....gttgatccctactctgtt 1637

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : N.Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1637	100.0	1637	21	AAA30922
2	1535	93.8	1230025	20	AAK91990
3	1484	90.7	273254	21	AAC81914
4	39	2.4	1038602	20	AAZ01425
5	23	1.4	43	21	AAA30923
6	22	1.3	786	22	AAI23777
7	21	1.3	2139	22	AAK30049
8	21	1.3	4673	22	AAK30048
9	21	1.3	5032	24	AAK54838

10	21	1.3	8220	16	AAK83529	P. falciparum Proj
11	21	1.3	8220	18	AAK72897	Plasmodium Proj3 g
12	21	1.3	8220	21	AAZ98286	P. falciparum Proj
13	21	1.3	19124	18	AAK72882	Plasmodium var-7 g
14	21	1.3	19124	21	AAZ98287	Plasmodium var-7 p
15	20	1.2	20	20	AAK95295	PCR primer used to
16	20	1.2	20	20	AAK95297	PCR primer used to
17	20	1.2	20	20	AAK95298	PCR primer used to
18	20	1.2	20	20	AAK92828	PCR primer used to
19	20	1.2	20	20	AAK92823	PCR primer used to
20	20	1.2	20	20	AAK92826	PCR primer used to
21	20	1.2	1770	22	AAK61373	A. thaliana chloro
22	20	1.2	2287	22	AAK77984	Nucleotide sequenc
23	20	1.2	33780	22	AAH24652	Human cDNA clone (
24	20	1.2	136284	24	AAK83575	Polynucleotide seq
25	19	1.2	562	22	AAH12882	Human neuroblastom
26	19	1.2	644	20	AAK20940	DNA encoding novel
27	19	1.2	764	22	AAI96504	Haemophilus influe
28	19	1.2	790	23	AAK74111	DNA sequence upstr
29	19	1.2	1001	22	AAK91438	Human prostate exp
30	19	1.2	1001	24	ABK37818	Human microglobuli
31	19	1.2	1362	23	ABV25656	Human glycosyl-pho
32	19	1.2	1604	24	ABA02319	Human cDNA sequenc
33	19	1.2	1689	24	ABL41351	Drosophila melanog
34	19	1.2	1833	22	AAH15049	Drosophila melanog
35	19	1.2	2611	23	ABL12876	Drosophila melanog
36	19	1.2	3052	23	ABL04386	DNA encoding SCAN
37	19	1.2	3752	24	ABK13001	Drosophila melanog
38	19	1.2	4507	23	ABL24540	Drosophila melanog
39	19	1.2	5586	23	ABL11410	Drosophila melanog
40	19	1.2	8034	23	ABL03008	Drosophila melanog
41	18	1.1	384	22	AAK23970	Human ovarian PCR-
42	18	1.1	384	24	ABK39720	cDNA encoding lung
43	18	1.1	396	22	AAK24335	Human ovarian PCR-
44	18	1.1	404	22	AAI01875	Human reproductive
45	18	1.1	404	23	ABL97168	Human testicular a
46	18	1.1	426	14	AAO60207	Human brain Expres
47	18	1.1	435	22	AAK30727	Human bone marrow
48	18	1.1	459	24	ABK97583	Human prostate spe
49	18	1.1	481	22	ABA54121	Human fetal liver
50	18	1.1	481	22	ABA23868	Probe #2334 for ge
51	18	1.1	481	22	AAK27835	Human bone marrow
52	18	1.1	481	22	AAI02320	Probe #2221 used t
53	18	1.1	484	22	AAH82924	Human ovarian tumo
54	18	1.1	486	22	AAH82533	Human ovarian tumo
55	18	1.1	687	21	AAH89558	Exo3-8 nucleotide
56	18	1.1	994	22	AAI06036	Human reproductive
57	18	1.1	994	23	ABL98601	Human testicular a
58	18	1.1	994	23	ABL98602	Human reproductive
59	18	1.1	995	23	ABL98602	Human testicular a
60	18	1.1	1298	22	AAH29860	C albicans apoptos
61	18	1.1	1681	22	AAK58306	Human GTP-binding
62	18	1.1	1876	22	AAK45074	Human secreted pro
63	18	1.1	1876	22	AAK33041	Human secreted pro
64	18	1.1	1876	24	ABO54209	Human ovarian anti
65	18	1.1	2001	22	AAK52772	Human polynucleoti
66	18	1.1	2043	22	AAK51788	Human polynucleoti
67	18	1.1	2677	22	AAI93897	Human stomach can
68	18	1.1	2677	22	AAI93897	Human stomach can
69	18	1.1	2727	23	ABV24810	Human prostate exp
70	18	1.1	2727	23	ABV25689	Human prostate exp
71	18	1.1	2748	22	AAI93884	Human stomach can
72	18	1.1	2748	22	AAH17695	Human cDNA sequenc
73	18	1.1	3022	20	AAZ20677	P. vulgaris chondr
74	18	1.1	3343	23	ABL20092	Drosophila melanog
75	18	1.1	3486	15	AAO71820	Chondroitinase ABC
76	18	1.1	3515	23	ABL05932	Drosophila melanog
77	18	1.1	3694	24	ABK83604	Human cDNA differe
78	18	1.1	3694	24	ABN95118	Gene #1616 used to
79	18	1.1	3699	24	ABK34876	Human cDNA encodin
80	18	1.1	3980	15	AAO74988	P. vulgaris chondr
81	18	1.1	3980	15	AAO74989	P. vulgaris chondr
82	18	1.1	3980	18	AAK56980	Proteus vulgaris c

83	18	1.1	3980	18	AAT58981	Mutated Proteus vu
84	18	1.1	5327	22	AA545357	Chemically pretrea
85	18	1.1	5327	24	ABK28190	DNA transcription
86	18	1.1	5591	23	ABL15092	Drosophila melanog
87	18	1.1	5591	23	ABL21058	Drosophila melanog
88	18	1.1	6065	22	ABL37262	Human musculoskele
89	18	1.1	6240	23	ABK69060	Human BRP genomic
90	18	1.1	6261	23	ABL14623	Drosophila melanog
91	18	1.1	6519	15	AA074990	P. vulgaris chondr
92	18	1.1	6519	18	AAT58982	Proteus vulgaris c
93	18	1.1	10200	23	ABL14622	Drosophila melanog
94	18	1.1	12768	23	ABL12988	Drosophila melanog
95	18	1.1	14736	19	AAV52304	Streptococcus pneu
96	18	1.1	16137	22	AAK87691	Human immune/haema
97	18	1.1	20445	24	AA519906	Reference sequence
98	18	1.1	28420	22	AAK78286	Human immune/haema
99	18	1.1	34488	22	AAF97854	Human neuroblastom
100	18	1.1	40862	24	ABL34073	Human immune syste

ALIGNMENTS

RESULT 1
AAA30922 standard; DNA; 1637 BP.

AAA30922;

22-SEP-2000 (first entry)

C. pneumoniae ATP/ADP translocase coding sequence.

ATP/ADP translocase; Chlamydia infection; diagnosis; therapy; ds.

Chlamydia pneumoniae.

Key Location/Qualifiers

FT CDS 51..1598

FT /tag= a

FT /product= ATP/ADP_translocase

XX MO200039157-A1.

XX 06-JUL-2000.

XX 22-DEC-1999; 99WO-CA01224.

XX 28-DEC-1998; 98US-0114060.

XX 12-MAR-1999; 99US-0123967.

XX 30-JUN-1999; 99US-0141271.

XX (CONN-) CONNAUGHT LAB LTD.

XX Murdin AD, Oomen RP, Wang J, Dunn P;

XX WPI: 2000-452368/39.

XX P-PSDB; AAY90265.

XX Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,

XX prevention and treatment of Chlamydia infection in mammals -

XX Claim 2; Fig 1; 81pp; English.

XX This sequence encodes the Chlamydia pneumoniae ATP/ADP translocase

XX of the invention. The protein, DNA encoding it, or a vaccine containing

XX the DNA or protein, are useful for diagnosing, preventing or treating

XX Chlamydia infection. The sequences can also be used in a method for

XX the detection of Chlamydia infection. Primers or probes derived from the

XX DNA sequence are useful in diagnostic tests for detecting Chlamydia

XX infection.

XX Sequence 1637 BP; 431 A; 327 C; 325 G; 554 T; 0 other;

Query Match 100.0%; Score 1637; DB 21; Length 1637;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAAATAAAAA	CTACAGATAGAAAAT	AAAGTATTTTCAGAGGGTAA	ATATGACAAAA	60
DB	1	GAAATAAAAA	CTACAGATAGAAAAT	AAAGTATTTTCAGAGGGTAA	ATATGACAAAA	60
QY	61	CCGAAGAAAA	CCCTTGGAAAAAT	TGGCGCTCTTCTGCGCGGAT	PACATATCAGAGC	120
DB	61	CCGAAGAAAA	CCCTTGGAAAAAT	TGGCGCTCTTCTGCGCGGAT	PACATATCAGAGC	120
QY	121	TAAAGAAAGT	TGCGCAATGTT	CCTAATGTTCTTGATTA	TCATTAACTATAGGT	180
DB	121	TAAAGAAAGT	TGCGCAATGTT	CCTAATGTTCTTGATTA	TCATTAACTATAGGT	180
QY	121	TAAAGAAAGT	TGCGCAATGTT	CCTAATGTTCTTGATTA	TCATTAACTATAGGT	180
DB	121	TAAAGAAAGT	TGCGCAATGTT	CCTAATGTTCTTGATTA	TCATTAACTATAGGT	180
QY	181	TACGGGAT	ACAAAGACACT	CTTATTTGGAGCTCTG	CTGTCGACAGGCAATAC	240
DB	181	TACGGGAT	ACAAAGACACT	CTTATTTGGAGCTCTG	CTGTCGACAGGCAATAC	240
QY	241	CTTTCATCAG	TTTGGCTGTGTG	CCCTGCTATATCTTAT	AGCTTATTTATGCA	300
DB	241	CTTTCATCAG	TTTGGCTGTGTG	CCCTGCTATATCTTAT	AGCTTATTTATGCA	300
QY	301	AGCTAAGTA	ATATTTAAGTA	AGAGCCTTATTTATG	CAGTGGACGCCCTTTTAA	360
DB	301	AGCTAAGTA	ATATTTAAGTA	AGAGCCTTATTTATG	CAGTGGACGCCCTTTTAA	360
QY	361	TTTTCTTTG	CCCTGTCCGAG	CTAATTTATCCGCTAC	GGCATTTTACATCTCAG	420
DB	361	TTTTCTTTG	CCCTGTCCGAG	CTAATTTATCCGCTAC	GGCATTTTACATCTCAG	420
QY	421	AATTTGCTAC	CGGTTTACAGG	CCATCTACCTCCAGAG	TTGCTAGACCTCGTGGCATCT	480
DB	421	AATTTGCTAC	CGGTTTACAGG	CCATCTACCTCCAGAG	TTGCTAGACCTCGTGGCATCT	480
QY	481	TAAGAACTG	ACATTTGCTG	CAATTTATGACTTGTG	CAACTATGGGAAACGTCATGC	540
DB	481	TAAGAACTG	ACATTTGCTG	CAATTTATGACTTGTG	CAACTATGGGAAACGTCATGC	540
QY	541	TATCTCTAT	ATGTTGCGGAT	TTGCTATGAAATTC	ACAAAATCCACGAAAGCGCTT	600
DB	541	TATCTCTAT	ATGTTGCGGAT	TTGCTATGAAATTC	ACAAAATCCACGAAAGCGCTT	600
QY	601	TCTACGCTT	TTTGGTATCGG	AGCTAATATTTCTT	ACTACCTCTGTCGCAATTG	660
DB	601	TCTACGCTT	TTTGGTATCGG	AGCTAATATTTCTT	ACTACCTCTGTCGCAATTG	660
QY	661	TTTGGGCTT	CAAGTTGAG	AGCTTCCGTTTCTG	AAAGGTGTAGATCCTTGGGAAATTTCTT	720
DB	661	TTTGGGCTT	CAAGTTGAG	AGCTTCCGTTTCTG	AAAGGTGTAGATCCTTGGGAAATTTCTT	720
QY	721	TACGCTTTT	GATGCTATG	ACTATTTGATCG	ACTTGTCTTATGCGCAATTTCTGCT	780
DB	721	TACGCTTTT	GATGCTATG	ACTATTTGATCG	ACTTGTCTTATGCGCAATTTCTGCT	780
QY	781	GGATCAAT	ATAGAGCTAT	TGACCGATCCGCT	GTATATATCCACAGAAATGCAAAAG	840
DB	781	GGATCAAT	ATAGAGCTAT	TGACCGATCCGCT	GTATATATCCACAGAAATGCAAAAG	840
QY	841	GGAAAAAG	GTGCTAAAC	CTTAAATGAAT	GATAGTAGCTTCTATCTTGTATAGAT	900
DB	841	GGAAAAAG	GTGCTAAAC	CTTAAATGAAT	GATAGTAGCTTCTATCTTGTATAGAT	900
QY	901	CTCTTATAT	CTTTTATTA	CTCTCTGTTAT	TGCTATGCTATGCTATTAACCTTAA	960
DB	901	CTCTTATAT	CTTTTATTA	CTCTCTGTTAT	TGCTATGCTATGCTATTAACCTTAA	960
QY	961	TGCAAGTAC	TTGGAAGT	AGCTGAAAGT	CAATGCAATATCCATATGTAATGTAATGTAATG	1020
DB	961	TGCAAGTAC	TTGGAAGT	AGCTGAAAGT	CAATGCAATATCCATATGTAATGTAATGTAATG	1020

```
OY 1021 AGTTCATGGGAACCTTCCTCTGAGCTGCGCTAGTATCCGTAATTATCATGCTATTG 1080
    |||
DB 1021 ATTTCATGGGAACCTTCCTCTCTGAGCTGCGCTAGTATCCGTAATTATCATGCTATTG 1080
OY 1081 TTGCTGTAAAGCTCATTTGTAATTTGGATGGTTAACTGAGAGCCCTAGTACTCTGTGA 1140
    |||
DB 1081 TTGCTGTAAAGCTCATTTGTAATTTGGATGGTTAACTGAGAGCCCTAGTACTCTGTGA 1140
OY 1141 TGGTCTCTTAACAGGTATGCTTTGCTGCTCTTCTTATCTTTAAGAACCAAGCTTCG 1200
    |||
DB 1141 TGGTCTCTTAACAGGTATGCTTTGCTGCTCTTCTTATCTTTAAGAACCAAGCTTCG 1200
OY 1201 GGCTGTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
    |||
DB 1201 GGCTGTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
OY 1261 AGAATATCTTTGGAATCCACAAATACGCTCTCTTGAAGTCACTAAAGAAAGGCT 1320
    |||
DB 1261 AGAATATCTTTGGAATCCACAAATACGCTCTCTTGAAGTCACTAAAGAAAGGCT 1320
OY 1321 ATATCCCTCTGACCAAGCAAAAGTCAAGGCTGCTATGATGATGCTGCTGCTGCTG 1380
    |||
DB 1321 ATATCCCTCTGACCAAGCAAAAGTCAAGGCTGCTATGATGATGCTGCTGCTGCTG 1380
OY 1381 CCCGCTTCGAAATCAGAGAGAGCTTTATCCACAAAGGTTGCTGCTTATCTGGA 1440
    |||
DB 1381 CCCGCTTCGAAATCAGAGAGAGCTTTATCCACAAAGGTTGCTGCTTATCTGGA 1440
OY 1441 GATATGAGCTATGAGCCCTTATCTTGCAGTATCTTCTTTCATCATTTGCTATTG 1500
    |||
DB 1441 GATATGAGCTATGAGCCCTTATCTTGCAGTATCTTCTTTCATCATTTGCTATTG 1500
OY 1501 TGGTCTCTGACCTAGTAAACAACTATCTTACGAGCTGCTGCTTAAAGCAAG 1560
    |||
DB 1501 TGGTCTCTGACCTAGTAAACAACTATCTTACGAGCTGCTGCTTAAAGCAAG 1560
OY 1561 AAGTGTCTAAGAGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
    |||
DB 1561 AAGTGTCTAAGAGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
OY 1621 GATCCCTACCTGCTTTT 1637
    |||
DB 1621 GATCCCTACCTGCTTTT 1637

RESULT 2
AA91990
ID AA91990 standard; DNA: 1230025 BP.
XX
AC AA91990;
XX
DT 13-SEP-1999 (first entry)
XX
DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO927105-A2.
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97RR-0014673.
XX
PA (GEST ) GENSET.
XX
PI Griffais R;
```

```
XX
DR WPI: 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Claim 1: Page 291-611; 1912pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC pneumoniae, and encodes proteins AA934584-Y35879. C. pneumoniae causes
CC respiratory disease such as pneumonia and bronchitis and is thought
CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC encoded by the open reading frames of the C. pneumoniae genome (see
CC AA934584-Y35879) can be used in immunogenic compositions as vaccines.
CC Vectors containing C. pneumoniae nucleotide sequences can also be
CC used as immunogenic compositions, especially where the vector directs
CC the expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 93.8%; Score 1535; DB 20; Length 1230025;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAATATAAAACATATGAGATAAATAAAGTATTTGAGAGGTAAATATGACAAAA 60
DB 404320 GAATATAAAACATATGAGATAAATAAAGTATTTGAGAGGTAAATATGACAAAA 404379
OY 61 CCGAAGAAAAACCTTTGGAAAAATGCGCTCTTCTTGGCCGATACATCTACGAGC 120
DB 404380 CCGAAGAAAAACCTTTGGAAAAATGCGCTCTTCTTGGCCGATACATCTACGAGC 404439
OY 121 TAAAGAAAGTCTGCCAATGTTCTTATGTTCTCTGATTAATTAATTAATGAGTGT 180
DB 404440 TAAAGAAAGTCTGCCAATGTTCTTATGTTCTCTGATTAATTAATTAATGAGTGT 404499
OY 181 TACGCATCAAAAGACACCTTATGTTGAGAGCTCCTGTTCTGTCGAGAGCAATAC 240
DB 404500 TACGCATCAAAAGACACCTTATGTTGAGAGCTCCTGTTCTGTTCTGTCGAGAGCAATAC 404559
OY 241 CTTTCATCAAGTTTGGCTGTTGTCCTGTCCTATATCTTATGCTTATTTATGCAA 300
DB 404560 CTTTCATCAAGTTTGGCTGTTGTCCTGTCCTATATCTTATGCTTATTTATGCAA 404619
OY 301 AGCTAAGTAAATATTTAAGTAGACAGCCCTATTTATGACAGTGGGAGCCCTTTTAA 360
DB 404620 AGCTAAGTAAATATTTAAGTAGACAGCCCTATTTATGACAGTGGGAGCCCTTTTAA 404679
OY 361 TTTTCTTGGCCCTGTTCCGAGCTGAATTATCCGCTACGCGATGTTTATACATCTACAG 420
DB 404680 TTTTCTTGGCCCTGTTCCGAGCTGAATTATCCGCTACGCGATGTTTATACATCTACAG 404739
OY 421 AATTGCTGACCGCTTACAGGCCATCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 480
DB 404740 AATTGCTGACCGCTTACAGGCCATCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 404799
OY 481 TAAGAAACAGGACATTTGCTGCAATTTATGACTTGTGCTAAGTGGGAGGCTATGC 540
DB 404800 TAAGAAACAGGACATTTGCTGCAATTTATGACTTGTGCTAAGTGGGAGGCTATGC 404859
OY 541 TATCTCTAATGTTCTGGGAGTTTGTATGAAATTCAAAAATCCAGAGCAACGCTT 600
DB 404860 TATCTCTAATGTTCTGGGAGTTTGTATGAAATTCAAAAATCCAGAGCAACGCTT 404919
OY 601 TCTACGCTCTTTTGGTATCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTTG 660
DB 404920 TCTACGCTCTTTTGGTATCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTTG 404979
OY 661 TTTGGGCTTCAAGAGTGAAGCTTCCGTTTCTAAGAGTGAAGTCTTGGGCAATTTCTT 720
DB 404980 TTTGGGCTTCAAGAGTGAAGCTTCCGTTTCTAAGAGTGAAGTCTTGGGCAATTTCTT 405039
OY 721 TACGCTTTTGAATGAGCTATGATGTAATCTGAGCTTGTCTTATGAGCAGTACTGGT 780
```

```

|||||
Db 405040 TACGCTTTTGAAGGCTATGACTATTGTACTGAGCTTGTCTTATGAGCAGTACTGGT 405099
QY 781 GGATCAATATAGAAACGATTGACGATCCTCGCTTCTATATCAGAAATSCAAAG 840
Db 405100 GGATCATATAGAAACGATTGACGATCCTCGCTTCTATATCAGAAATSCAAAG 405159
QY 841 GGAATAAAGGCTCTAAACCTAAATATGAAATGAAAGATAGCTTCTATCTTATGAT 900
Db 405160 GGAATAAAGGCTCTAAACCTAAATATGAAATGAAAGATAGCTTCTATCTTATGAT 405219
QY 901 CTCCTATATCTTCTTATATCTCTCTGTTATCTGTTATCTGATTTGATTTAA 960
Db 405220 CTCCTATATCTTCTTATATCTCTCTGTTATCTGTTATCTGATTTGATTTAA 405279
QY 961 TCGAAGTACTTGAATAAGCTCAAGCTCAATATCTATATGATGATGATAGT 1020
Db 405280 TCGAAGTACTTGAATAAGCTCAAGCTCAATATCTATATGATGATGATAGT 405339
QY 1021 ACTTCATGGGGAACTTCTCTCTCTGAGCTGGCGTAGTATCCGTAATCATGTA 1080
Db 405340 ACTTCATGGGGAACTTCTCTCTCTGAGCTGGCGTAGTATCCGTAATCATGTA 405399
QY 1081 TTGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 405400 TTGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405459
QY 1141 TGGTCTCTCTAAGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 405460 TGGTCTCTCTAAGAGTATGATGATGATGATGATGATGATGATGATGAT 405519
QY 1201 GCGTGTGCTTATGTTGCTGATACACTCTCTATGCTAGCTGTTGTTGGA 1260
Db 405520 GCGTGTGCTTATGTTGCTGATACACTCTCTATGCTAGCTGTTGTTGGA 405579
QY 1261 AGAATATCTCTGGAATATCACAATAATAGCTCTCTGCTACTCACTAAGA 1320
Db 405580 AGAATATCTCTGGAATATCACAATAATAGCTCTCTGCTACTCACTAAGA 405639
QY 1321 ATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGTAAAGGCTATATGAT 1380
Db 405640 ATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGTAAAGGCTATATGAT 405699
QY 1381 CCGCGTGGGAATATCAGAGAGAGCTTATATCCACAAAGGTTGCTGCT 1440
Db 405700 CCGCGTGGGAATATCAGAGAGAGCTTATATCCACAAAGGTTGCTGCT 405759
QY 1441 GATTGGAGCTATGACCCCTTATCTTGAGTATCTTCTTCAATGCTATTTG 1500
Db 405760 GATTGGAGCTATGACCCCTTATCTTGAGTATCTTCTTCAATGCTATTTG 405819
QY 1501 TGGTCTCTGCAACTAAGTTAAACAACTATCTTACGCGAGTCTGCTTAA 1560
Db 405820 TGGTCTCTGCAACTAAGTTAAACAACTATCTTACGCGAGTCTGCTTAA 405879
QY 1561 AAGTGGCTCAAGAAATTCAGTCTGCTTCTTCAATAGTGTCTTCTTACT 1620
Db 405880 AAGTGGCTCAAGAAATTCAGTCTGCTTCTTCAATAGTGTCTTCTTACT 405939
QY 1621 GATCCCTACTGCTTTT 1637
Db 405940 GATCCCTACTGCTTTT 405956

```

RESULT 3
AAC81914/c
ID AAC81914 standard; DNA: 273254 bp.

AC AAC81914;
XX
XX 27-FEB-2001 (first entry)
XX
DE Chlamydia pneumoniae genome DNA.

```

XX Genome; diagnosis; vaccine; ds.
KW Chlamydia pneumoniae.
OS WO200027994-A2.
PN 18-MAY-2000.
XX 12-NOV-1999; 99WO-US26923.
PF 12-NOV-1999; 99US-0108279.
PR 08-APR-1999; 99US-0128606.
XX (REGC ) UNIV CALIFORNIA.
PA Stephens R, Mitchell W, Kalman S, Davis R;
PI WPI: 2000-376516/32.
DR
XX Isolated nucleic acid for use in diagnostic and analytical methods
PT encodes genomic sequence of Chlamydia pneumoniae
PR
XX Claim 2: Page 128-320; 320pp; English.
XX
XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC pneumoniae protein (P1), given in the specification. The isolated nucleic
CC acid is useful for diagnostic and analytical methods, such as,
CC hybridization-based assays or amplification-based assays. The protein may
CC be used for diagnostic purposes, for their enzymatic or structural
CC activity, or as a vaccine. The invention also describes (1) a probe
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC cassette comprising N1 under the transcriptional regulation of a
CC transcriptional termination region; (4) a cell comprising an expression
CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell, and the cellular progeny of the host cell;
CC (5) a method for producing a P1 comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other
CC proteins; (6) a purified polypeptide composition comprising at least 50
CC weight % of P1; and (7) a monoclonal antibody binding specifically to the
CC peptide of (6).
XX
SQ Sequence 273254 bp; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;
Query Match 90.7%; Score 1484; DB 21; Length 273254;
Best Local Similarity 99.8%; Pred No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAAATATAAAATATCAGAAATAGAAATTAATTTTCAGAGGTTAAATATGACAAAA 60
Db 212470 GAAATATAAAATATCAGAAATAGAAATTAATTTTCAGAGGTTAAATATGACAAAA 212411
QY 61 CCGAAGAAAAACCTTTTGGAAAAATTCGCTCTTCTTGTGGCCGATACACTACGAGC 120
Db 212410 CCGAAGAAAAACCTTTTGGAAAAATTCGCTCTTCTTGTGGCCGATACACTACGAGC 212351
QY 121 TAAAGAAAGTCTGCCAATGTTCTTAATGTTCTGATTTACATTTAATACGAGGT 180
Db 212350 TAAAGAAAGTCTGCCAATGTTCTTAATGTTCTGATTTACATTTAATACGAGGT 212291
QY 181 TACGCAATCAAAAGACACTTTATTTGGAGAGCTCTGTTCTGCTCAGAGCAATAC 240
Db 212290 TACGCAATCAAAAGACACTTTATTTGGAGAGCTCTGTTCTGCTCAGAGCAATAC 212231
QY 241 CTTTCATCAAGTTTGGCTTGTTCCTCTGCTATTAATCTTATGCTTATTTATGCAA 300
Db 212230 CTTTCATCAAGTTTGGCTTGTTCCTCTGCTATTAATCTTATGCTTATTTATGCAA 212171
QY 301 AGCTAAGTAATATTTTAAAGTAAAGAGGCTTATTTATGCACTGGAGCCCTTTTAA 360

```



```

Db 212170 AGCTAGTAATATTTTAAGTAAGCAGGCGCTTATTTTATGCACTGGGAAAGCCCTTTTAA 212111
OY 361 TTTCTTTGCCCTGTTCGCGACTAATTTATTCGCTACGCGATGTTTATCTTACG 420
    |||||||
Db 212110 TTTCTTTGCCCTGTTCGCGACTAATTTATTCGCTACGCGATGTTTATCTTACG 212051
OY 421 AATTTGCTGACCGTTTACAGGCGATCTACCTCAGAGATGCTAGAGACTCGTGGCATCT 480
    |||||||
Db 212050 AATTTGCTGACCGTTTACAGGCGATCTACCTCAGAGATGCTAGAGACTCGTGGCATCT 211991
OY 481 TAAGAACTGCGACTTTGCTGATTTTATGTACTGCTGAGACTATGCGGAGCGTCATGC 540
    |||||||
Db 211990 TAAGAACTGCGACTTTGCTGATTTTATGTACTGCTGAGACTATGCGGAGCGTCATGC 211931
OY 541 TATCTATATGTTGGGGATTTGCTAATGAATACAAAATTCACGAGCAAGACCGTT 600
    |||||||
Db 211930 TATCTATATGTTGGGGATTTGCTAATGAATACAAAATTCACGAGCAAGACCGTT 211871
OY 601 TCTACGCTCTTTTCGCTACGAGCTAATATTTCTTACTAGCTCTGCTGCGCATTTG 660
    |||||||
Db 211870 TCTACGCTCTTTTCGCTACGAGCTAATATTTCTTACTAGCTCTGCTGCGCATTTG 211811
OY 661 TTTGGGCTTCAAAAGTTGAGAGCTCCGTTTCTGAAGGTGTAGATCCCTGGGGAATTTCTT 720
    |||||||
Db 211810 TTTGGGCTTCAAAAGTTGAGAGCTCCGTTTCTGAAGGTGTAGATCCCTGGGGAATTTCTT 211751
OY 721 TACGCTCTTTGATGGCTATGACTATTTATCTGAGACTTGTCTTATGGCACTTACTGCT 780
    |||||||
Db 211750 TACGCTCTTTGATGGCTATGACTATTTATCTGAGACTTGTCTTATGGCACTTACTGCT 211691
OY 781 GGATCAATAGAAGCTATTTGACCGATCCGCTTCTATATCTCAAGAAAGATGCAAAAG 840
    |||||||
Db 211690 GGATCAATAGAAGCTATTTGACCGATCCGCTTCTATATCTCAAGAAAGATGCAAAAG 211631
OY 841 GGAAGAAAGGTCTAAGCTAATATGAATGAAGATAGCTTCTTATCTTGAATGAT 900
    |||||||
Db 211630 GGAAGAAAGGTCTAAGCTAATATGAATGAAGATAGCTTCTTATCTTGAATGAT 211571
OY 901 CTCCTTATATCTTTTATTAATCTCTGTGATATGCGATGCGATGTTTGCATTAATCTAA 960
    |||||||
Db 211570 CTCCTTATATCTTTTATTAATCTCTGTGATATGCGATGCGATGTTTGCATTAATCTAA 211511
OY 961 TCGAAGTGACTTGGAAAAGTCAAGTCAAACTCAATATCTCAATATGTAATGACTATAGTG 1020
    |||||||
Db 211510 TCGAAGTGACTTGGAAAAGTCAAGTCAAACTCAATATCTCAATATGTAATGACTATAGTG 211451
OY 1021 AGTTCATGGGGAATCTTCTCTCTGAGCTGGCTAGTATCCGTACTTATCATGCTATTGG 1080
    |||||||
Db 211450 AGTTCATGGGGAATCTTCTCTCTGAGCTGGCTAGTATCCGTACTTATCATGCTATTGG 211391
OY 1081 TTGGTGTGAACGTCATTTGCTAATTTGATGTTAAGTGAAGCCCTAGTCACTCCTGTCA 1140
    |||||||
Db 211390 TTGGTGTGAACGTCATTTGCTAATTTGATGTTAAGTGAAGCCCTAGTCACTCCTGTCA 211331
OY 1141 TGGTTCCTCTTAACAGGTATCGTTTCTTCGCTCTGTTATCTTGAAGCAAGCTTCTG 1200
    |||||||
Db 211330 TGGTTCCTCTTAACAGGTATCGTTTCTTCGCTCTGTTATCTTGAAGCAAGCTTCTG 211271
OY 1201 GCGTGTGCTATGTTGGGTACAACTCCTCTCATCTAGCTGTGTTGTGCGAGGTATAC 1260
    |||||||
Db 211270 GCGTGTGCTATGTTGGGTACAACTCCTCTCATCTAGCTGTGTTGTGCGAGGTATAC 211211
OY 1261 AGAATATCTTTTTCGAAAATTCACAAAATAGCGTCTTTGAGTCACTAAAGAAATGGCGT 1320
    |||||||
Db 211210 AGAATATCTTTTTCGAAAATTCACAAAATAGCGTCTTTGAGTCACTAAAGAAATGGCGT 211151
OY 1321 ATATCCCTCTTTCGAAAAGTCAAAAGTCAAGGTCAAGGCTCTATGATGTAAGTTGGCG 1380
    |||||||
Db 211150 ATATCCCTCTTTCGAAAAGTCAAAAGTCAAGGTCAAGGCTCTATGATGTAAGTTGGCG 211091
OY 1381 CCCGCTTGGAAAATTCAGAGAGGCTTTATATCAACAAGGTTGCTGTTATCTGTGAAA 1440
    |||||||
Db 211090 CCCGCTTGGAAAATTCAGAGAGGCTTTATATCAACAAGGTTGCTGTTATCTGTGAAA 211031

```

```

OY 1441 GATTTGAGCTATGACCCCTTATCTTGCAGTATCTCTTTTCAATGCTATTGGT 1500
    |||||||
Db 211030 GATTTGAGCTATGACCCCTTATCTTGCAGTATCTCTTTTCAATGCTATTGGT 210971
OY 1501 TGGTTCCTGCAACTAAGTTAAACAACATATCTTATGCGCAGTCTGCTTTAAACAAG 1560
    |||||||
Db 210970 TGGTTCCTGCAACTAAGTTAAACAACATATCTTATGCGCAGTCTGCTTTAAACAAG 210911
OY 1561 AAGTGGCTTCAGAACATTCAGCTCTGCTCTTTCATPAGAGTTGCTTCTTACTTGTG 1620
    |||||||
Db 210910 AAGTGGCTTCAGAACATTCAGCTCTGCTCTTTCATPAGAGTTGCTTCTTACTTGTG 210851
OY 1621 GATCCCTACCTGCTTTT 1637
    |||||||
Db 210850 GATCCCTACCTGCTTTT 210834

RESULT 4
AAZ01425
ID AAZ01425 standard; DNA; 1038602 BP.
XX
AC AAZ01425;
XX
DT 07-OCT-1999 (first entry)
XX
DE Complete genome sequence of Chlamydia trachomatis.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perithelpatitis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS Chlamydia trachomatis.
XX
W09928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-1B01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GBST ) GENSET.
PI Griffais R;
XX
DR WPI: 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1: Page 373-656; 1755pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AAY36754-737949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal urethritis,
CC epididymitis, cervicitis, salpingitis, perithelpatitis, Bartholinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
SO Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;
Query Match 2.4%; Score 39; DB 20; Length 1038602;
Best Local Similarity 100.0%; Pred. No. 5.7e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 123 AAGAAAGTCTGCAATGTTCTTAAGTTCTTGTATT 161
 |||
 DB 392479 AAGAAAGTCTGCAATGTTCTTAAGTTCTTGTATT 392517

RESULT 5

AAA30923
 ID AAA30923 standard; DNA; 43 BP.

AC AAA30923;

DT 22-SEP-2000 (first entry)

DE PCR primer for C. pneumoniae ATP/ADP translocase coding sequence.

KM ATP/ADP translocase; Chlamydia infection; diagnosis; therapy;
 KW PCR primer; ss.

OS Chlamydia pneumoniae.

PN WO200039157-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-CA01224.

PR 28-DEC-1998; 98US-0114060.

PR 12-MAR-1999; 99US-0123967.

PR 30-JUN-1999; 99US-0141271.

PA (CONN-) CONNAUGHT LAB LTD.

PI Murdin AD, Oomen RP, Wang J, Dunn P;

DR WPI; 2000-452368/39.

PT Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,
 prevention and treatment of Chlamydia infection in mammals -

PS Claim 32; Page 48; 81pp; English.

CC This sequence represents a PCR primer used to isolate DNA encoding
 the Chlamydia pneumoniae ATP/ADP translocase of the invention. The
 CC protein, DNA encoding it, or a vaccine containing the DNA or protein, are
 CC useful for diagnosing, preventing or treating Chlamydia infection. The
 CC sequences can also be used in a method for the detection of Chlamydia
 CC infection. Primers or probes derived from the DNA sequence are useful in
 CC diagnostic tests for detecting Chlamydia infection.

SO Sequence 43 BP; 20 A; 12 C; 8 G; 3 T; 0 other;

Query Match 1.4%; Score 23; DB 21; Length 43;

Best Local Similarity 100.0%; Pred. No. 0.57; Mismatches 0; Indels 0; Gaps 0;

QY 51 ATGACAAAACCGAAGAAAAC 73
 |||

DB 21 ATGACAAAACCGAAGAAAAC 43

RESULT 6

AAAL23777
 ID AAAL23777 standard; cDNA; 786 BP.

AC AAAL23777;

DT 07-DEC-2001 (first entry)

DE Human breast cancer expressed polynucleotide 16234.

KW Human; breast cancer; cell marker; cytostatic; ss.

OS Homo sapiens.

PN WO200151628-A2.

PD 19-JUL-2001.

PF 10-JAN-2001; 2001WO-US00798.

PR 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

PR 24-MAR-2000; 2000US-0192099.

PR 29-MAR-2000; 2000US-0193480.

PR 15-MAY-2000; 2000US-0205230.

PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Lillie J, Xu Y, Wang Y, Steinmann K;

DR WPI; 2001-451856/48.

PT New peptide useful as a marker for the diagnosis of breast cancer -

PS Claim 1; Page 2966; 3695pp; English.

CC The invention relates to human breast cancer expressed polynucleotides
 CC (AAI07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterizing treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.

SO Sequence 786 BP; 201 A; 177 C; 169 G; 238 T; 1 other;

Query Match 1.3%; Score 22; DB 22; Length 786;

Best Local Similarity 100.0%; Pred. No. 1.7; Mismatches 0; Indels 0; Gaps 0;

QY 908 TATTTCTTTTATTACTCTCTTG 929
 |||

DB 485 TATTTCTTTTATTACTCTCTTG 506

RESULT 7

AAS30049
 ID AAS30049 standard; DNA; 2139 BP.

AC AAS30049;

DT 21-NOV-2001 (first entry)

DE Human lung antigen genomic DNA #119.

KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
 KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
 KW antirheumatic; antiproliferative; cytostatic; cardiact; neuroprotective;
 KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
 KW ophtalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
 KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
 KW cerebrovascular disorder; nervous system disorder; bacterial infection;
 KW fungal infection; viral infection; ocular disorder; endocrine disorder;
 KW gastrointestinal disorder; renal disorder; respiratory disorder;
 KW wound healing; skin aging; organ transplantation; food preservative;
 KW tissue regeneration; anti-infertility; food additive.

OS Homo sapiens.

PN WO200155303-A2.

PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01301.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.

XX Isolated polypeptide for treating, preventing and/or prognosing
PT respiratory disorders related to the lung including lung cancers and
PT also for testing and detection e.g. diagnosis -

PS Claim 1; SEQ ID NO 313; 507pp; English.

CC Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode
CC the lung antigen polypeptides of the invention. Lung antigen polypeptides
CC and their associated polynucleotides are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by detecting the presence or
CC absence of a mutation in a lung antigen polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Query Match 1.3%; Score 21; DB 22; Length 2139;

Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 700 TAGATCCTTGGGGAATTCTT 720
|||||

Db 31 TAGATCCTTGGGGAATTCTT 51

RESULT 8

AAS30048
ID AAS30048 standard; DNA; 4673 BP.

AC AAS30048;

DT 21-NOV-2001 (first entry)

DE Human lung antigen genomic DNA #118.

XX Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
KM chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KM antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KM cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KM ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KM hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KM cerebrovascular disorder; nervous system disorder; bacterial infection;
KM fungal infection; viral infection; ocular disorder; endocrine disorder;
KM gastrointestinal disorder; renal disorder; respiratory disorder;
KM wound healing; skin aging; organ transplantation; food preservative;
KM tissue regeneration; anti-infertility; food additive.

OS Homo sapiens.

XX WO200155303-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01301.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0211486.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232968.
PR 14-SEP-2000; 2000US-0232969.
PR 14-SEP-2000; 2000US-0232996.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233403.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

XX	PR	02-OCT-2000	2000US-0237040.
XX	PR	13-OCT-2000	2000US-0238933.
XX	PR	20-OCT-2000	2000US-0240960.
XX	PR	20-OCT-2000	2000US-0241221.
XX	PR	20-OCT-2000	2000US-0241785.
XX	PR	20-OCT-2000	2000US-0241786.
XX	PR	20-OCT-2000	2000US-0241787.
XX	PR	20-OCT-2000	2000US-0241808.
XX	PR	20-OCT-2000	2000US-0241809.
XX	PR	01-NOV-2000	2000US-0244826.
XX	PR	08-NOV-2000	2000US-0244617.
XX	PR	08-NOV-2000	2000US-0246475.
XX	PR	08-NOV-2000	2000US-0246476.
XX	PR	08-NOV-2000	2000US-0246477.
XX	PR	08-NOV-2000	2000US-0246478.
XX	PR	08-NOV-2000	2000US-0246523.
XX	PR	08-NOV-2000	2000US-0246524.
XX	PR	08-NOV-2000	2000US-0246525.
XX	PR	08-NOV-2000	2000US-0246526.
XX	PR	08-NOV-2000	2000US-0246527.
XX	PR	08-NOV-2000	2000US-0246528.
XX	PR	08-NOV-2000	2000US-0246532.
XX	PR	08-NOV-2000	2000US-0246609.
XX	PR	08-NOV-2000	2000US-0246610.
XX	PR	08-NOV-2000	2000US-0246611.
XX	PR	08-NOV-2000	2000US-0246613.
XX	PR	17-NOV-2000	2000US-0249207.
XX	PR	17-NOV-2000	2000US-0249208.
XX	PR	17-NOV-2000	2000US-0249209.
XX	PR	17-NOV-2000	2000US-0249210.
XX	PR	17-NOV-2000	2000US-0249211.
XX	PR	17-NOV-2000	2000US-0249212.
XX	PR	17-NOV-2000	2000US-0249213.
XX	PR	17-NOV-2000	2000US-0249214.
XX	PR	17-NOV-2000	2000US-0249215.
XX	PR	17-NOV-2000	2000US-0249216.
XX	PR	17-NOV-2000	2000US-0249217.
XX	PR	17-NOV-2000	2000US-0249218.
XX	PR	17-NOV-2000	2000US-0249244.
XX	PR	17-NOV-2000	2000US-0249245.
XX	PR	17-NOV-2000	2000US-0249264.
XX	PR	17-NOV-2000	2000US-0249265.
XX	PR	17-NOV-2000	2000US-0249297.
XX	PR	17-NOV-2000	2000US-0249299.
XX	PR	17-NOV-2000	2000US-0249300.
XX	PR	01-DEC-2000	2000US-0250160.
XX	PR	01-DEC-2000	2000US-0250391.
XX	PR	05-DEC-2000	2000US-0251030.
XX	PR	05-DEC-2000	2000US-0251988.
XX	PR	05-DEC-2000	2000US-0256719.
XX	PR	06-DEC-2000	2000US-0251479.
XX	PR	08-DEC-2000	2000US-0251856.
XX	PR	08-DEC-2000	2000US-0251866.
XX	PR	08-DEC-2000	2000US-0251869.
XX	PR	08-DEC-2000	2000US-0251989.
XX	PR	11-DEC-2000	2000US-0254097.
XX	PR	05-JAN-2001	2001US-0255678.
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Rosen CA, Barash SC, Ruben SM;	
XX	DR	WPI; 2001-457723/49.	
XX	PT	Isolated polypeptide for treating, preventing and/or prognosing	
XX	PT	respiratory disorders related to the lung including lung cancers and	
XX	PT	also for testing and detection e.g. diagnosis -	
XX	XX	Claim 1; SEQ ID No 312; 507pp; English.	

CC	Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode
CC	the lung antigen polypeptides of the invention. Lung antigen polypeptides
CC	and their associated polynucleotides are useful in the diagnosis,
CC	treatment and prevention of various types of disorders in e.g. humans,
CC	mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC	pathological condition can be determined by detecting the presence or
CC	absence of a mutation in a lung antigen polynucleotide. The treatable
CC	disorders include autoimmune diseases such as rheumatoid arthritis,
CC	hyperproliferative disorders such as neoplasms of the breast or liver,
CC	cardiovascular disorders such as cardiac arrest, cerebrovascular
CC	disorders such as cerebral ischaemia, nervous system disorders such as
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC	ocular disorders such as corneal infection, endocrine disorders such as
CC	premature labour and infertility, gastrointestinal disorders such as
CC	Crohn's disease, renal disorders such as glomerulonephritis and
CC	respiratory disorders such as asthma and pleurisy. The polypeptides can
CC	also be used to aid wound healing, to prevent skin aging due to sunburn,
CC	to maintain organs before transplantation, to regenerate tissues and in
CC	chemotaxis. The polypeptides can also be used as a food additive or
CC	preservative to increase or decrease storage capabilities.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences .
CC	
CC	Query Match 1.3%; Score 21; DB 22; Length 4673;
CC	Best Local Similarity 100.0%; Pred. No. 5.3; 0; Indels 0; Gaps 0;
CC	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC	
CC	Dy 700 TAGATCCTTGGGGAATTTCTT 720
CC	
CC	Db 2565 TAGATCCTTGGGGAATTTCTT 2585
CC	
CC	RESULT 9
CC	AAS94838
CC	ID AAS94838 standard; DNA: 5032 BP.
CC	XX AAS94838;
CC	XX
CC	DT 14-FEB-2002 (first entry)
CC	XX
CC	DE Human DNA sequence #93 expressed during foam cell differentiation.
CC	XX
CC	KW Human: foam cell differentiation; atherosclerosis; cerebral stroke;
CC	KW cardiovascular disorder; coronary artery disease; gene therapy. ds.
CC	XX
CC	OS Homo sapiens.
CC	XX
CC	PN WO200177389-A2.
CC	XX
CC	PD 18-OCT-2001.
CC	XX
CC	PE 04-APR-2001; 2001WO-US11128.
CC	XX
CC	XX 05-APR-2000; 2000US-195106P.
CC	XX
CC	PA (INCYT) INCYTE GENOMICS INC.
CC	XX
CC	PI Shiftman D, Somogyi R, Lawn R, Sellhammer JJ, Porter GJ, Mikita T;
CC	PI Tai J;
CC	DR WPI; 2002-010925/01.
CC	XX
CC	PT Composition useful for diagnosis of conditions, disorders or diseases
CC	PT associated with atherosclerosis, comprises several polynucleotides that
CC	PT are differentially expressed in foam cell development -
CC	XX
CC	PS Claim 1; Page 146-147; 315pp; English.
CC	XX
CC	CC The present invention relates to the isolation of human polynucleotide
CC	CC sequences that are differentially expressed during foam cell
CC	CC differentiation. The polynucleotide sequences of the invention or a
CC	CC composition comprising these polynucleotides are useful as a thick

CC throughput method for detecting altered expression of one or more
 CC polynucleotides in a sample. The polynucleotides can be used in the
 CC diagnosis of disorders associated with foam cell development such as
 CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
 CC coronary artery disease. The polynucleotide sequences can also be used
 CC as PCR primers and probes. The polynucleotides of the invention are also
 CC useful in gene therapy. AAS94746-AAS95021 represent the human
 CC polynucleotide sequences of the invention which are differentially
 CC expressed during foam cell differentiation.

SO Sequence 5032 BP; 1473 A; 1026 C; 1055 G; 1477 T; 1 other;

Query Match 1.3%; Score 21; DB 24; Length 5032;
 Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 TAGATCCTTGGGAATTCTT 720
 Db 2927 TAGATCCTTGGGAATTCTT 2947

RESULT 10

AA083529
 ID AA083529 standard; DNA; 8220 BP.

AC AA083529;

DT 22-SEP-1995 (first entry)

DE P. falciparum Proj3 gene.

XX Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;

KW vaccine; ss.

OS Plasmodium falciparum.

PN W09507353-A.

PD 16-MAR-1995.

PF 07-SEP-1994; 94WO-US10230.

PR 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

PI Wellens TE;

DR WPI: 1995-123427/16.

DR P-PSDB; AAR70236.

PT New erythrocyte binding domain polypeptide(s) - isolated from

PT Plasmodium binding proteins, used in diagnosis, treatment and

PT prevention of malaria

PS Disclosure: Page 61-65; 81pp; English.

XX Erythrocyte binding ligand (EBL) family genes were cloned from

CC P. falciparum chromosome 7 subsegment libraries constructed during

CC genetic studies of the chloroquine resistance locus. The 4 genes,

CC EBL-e1 (AA083526), E31a (AA083527), EBL-e2 (AA083528) and Proj3

CC (AA083529), encode the proteins given in AAR70233-36, respectively.

CC The binding domains of such proteins can be expressed e.g. in E. coli,

CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected

CC cells, and provide protection against P. falciparum.

SO Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T; 0 other;

Query Match 1.3%; Score 21; DB 16; Length 8220;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAACCTTTGGAAAT 84
 Db 5046 AAGAAAACCTTTGGAAAT 5066

RESULT 11

AAT72897
 ID AAT72897 standard; cDNA; 8220 BP.

AC AAT72897;

DT 07-OCT-1997 (first entry)

DE Plasmodium Proj3 gene.

XX DBL gene family; SAbP: stalic acid binding protein; vaccine; therapy;

KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;

KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;

XX Plasmodium; ss.

OS Plasmodium falciparum.

PN W09640766-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US09508.

PR 07-JUN-1995; 95US-0487826.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

PI Wellens TE;

DR WPI: 1997-052231/05.

DR P-PSDB; AAM22482.

PT New malaria vaccines - contains cysteine-rich DBL family protein

PT binding domains homologous domains of the Duffy and stalic acid

PT binding proteins

PS Disclosure: Page 48-50; 96pp; English.

XX This sequence represents the Proj3 gene of Plasmodium. Proj3 belongs to

CC the Duffy binding like (DBL) family of genes which have homology to the

CC Duffy antigen binding protein (DABP) and stalic acid binding protein

CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The

CC var family of genes modulate cytoadherence and antigenic variation of

CC Plasmodium infected erythrocytes. SAbP and the Duffy antigen binding

CC protein (DABP) are soluble proteins that appear in the culture

CC supernatant after infected erythrocytes release merozoites. DABP and

CC SAbP mediate the binding of merozoites and schizonts to the erythrocyte

CC surface. These proteins are necessary for erythrocyte invasion by the

CC parasite. This sequence can be used in the compositions of the

CC invention. The compositions are for the treatment and prevention of

CC malaria, and comprise either a nucleotide sequence or encoded polypeptide

CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a

CC family of genes having homology with conserved regions of DABP and SAbP.

CC The compositions are used for the treatment and prevention of malaria.

CC They are also used in the preparation of vaccines for inducing a

CC protective immune response in a mammal to Plasmodium merozoites

CC (especially Plasmodium falciparum or Plasmodium vivax).

SO Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T; 0 other;

Query Match 1.3%; Score 21; DB 18; Length 8220;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AAGAAAACCTTTGGAAAT 84
 |||||||||||||||||||
 Db 5046 AAGAAAACCTTTGGAAAT 5066

RESULT 12
 AA298286
 ID AA298286 standard; DNA; 8220 BP.
 XX
 AC AA298286;
 XX
 DT 13-JUN-2000 (first entry)
 XX
 DE P. falciparum Proj3 binding domain polypeptide encoding DNA.
 XX
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KW DABP; Stalic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
 KW protozoacide; ds.
 XX
 OS Plasmodium falciparum.
 XX
 PN US5993827-A.
 XX
 PD 30-NOV-1999.
 XX
 PF 07-JUN-1995; 95US-0487826.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
 XX
 DR WPI; 2000-194198/17.
 XX
 DR P-PSDB; AAT77904.
 XX
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX
 PS Disclosure; Columns 71-80; 93pp; English.
 XX
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria. The present sequence represents the Proj3 binding domain
 CC polypeptide encoding DNA.
 XX
 SQ Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T; 0 other;

Query Match 1.3%; Score 21; DB 21; Length 8220;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AAGAAAACCTTTGGAAAT 84
 |||||||||||||||||||
 Db 5046 AAGAAAACCTTTGGAAAT 5066

RESULT 13
 AAT72882
 ID AAT72882 standard; CDNA; 19124 BP.
 XX
 AC AAT72882;

XX
 DT 12-SEP-1997 (first entry)
 XX
 DE Plasmodium var-7 gene.
 XX
 KW DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium; ss.
 XX
 OS Plasmodium vivax.
 XX
 PN Plasmodium falciparum.
 XX
 FH Key location/Qualifiers
 FT exon 7317..15139
 FT /*tag= a
 FT /number= 1
 FT 15140..16205
 FT /*tag= b
 FT /number= 1
 FT 16206..17552
 FT exon 16206..17552
 FT /*tag= c
 FT /number= 2
 FT /note= "no stop codon given"
 XX
 PN WO9640766-A2.
 XX
 PD 19-DEC-1996.
 XX
 PF 07-JUN-1996; 96WO-US09508.
 XX
 PR 07-JUN-1995; 95US-0487826.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 XX
 DR Wellens TE;
 XX
 DR WPI; 1997-052231/05.
 XX
 DR P-PSDB; AAW2475.
 XX
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stalic acid
 PT binding proteins
 XX
 PS Claim 4; Page 56-61; 96pp; English.
 XX
 CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stalic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions can be used for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family. a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).
 XX
 SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 1.3%; Score 21; DB 18; Length 19124;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY      64  AAGAAAACCTTTGGAAT  84
      |||||||
DB      12366  AAGAAAACCTTTGGAAT  12386

RESULT 14
AA298287
ID      AA298287 standard; DNA: 19124 BP.
XX
XX      AA298287;
XX
XX      13-JUN-2000 (first entry)
XX
XX      Plasmodium var-7 polypeptide encoding DNA.
DE
XX      DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW      DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
KW      protozoa; var-7; ds.
XX
XX      Plasmodium sp.
OS
XX      US593827-A.
XX      30-NOV-1999.
PD
XX      07-JUN-1995; 95US-0487826.
XX
XX      10-SEP-1993; 93US-0119677.
PR
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX      Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
PI      WPI: 2000-194198/17.
XX      P-PSDB; AAY77904.
DR
XX      Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT      falciparum erythrocyte binding proteins useful for vaccinating against
PT      malaria -
XX
XX      Disclosure; Columns 91-108; 93pp; English.
XX
XX      The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC      (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC      identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC      Binding Protein (SABP), which are soluble proteins that appear in the
CC      culture supernatant after erythrocytes infected with malaria release
CC      merozoites. Immunochemical studies indicate that DABP and SABP are the
CC      respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC      and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC      used to vaccinate against malaria, especially caused by P. falciparum.
CC      Immunization with the polypeptide provides effective protection against
CC      malaria. The present sequence represents the var-7 polypeptide encoding
CC      DNA.
XX
XX      Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;
SQ
Query Match      1.3%; Score 21; DB 21; Length 19124;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      64  AAGAAAACCTTTGGAAT  84
      |||||||
DB      12366  AAGAAAACCTTTGGAAT  12386

RESULT 15
AAK95295/C
ID      AAK95295 standard; DNA: 20 BP.
XX
XX      AAK95295;
XX
XX      13-SEP-1999 (first entry)
XX

```

```

XX      PCR primer used to amplify an ORF of Chlamydia pneumoniae.
DE
XX      Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW      sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW      vaccine; neutralising epitope; PCR primer; ss.
XX
XX      Synthetic.
OS      Chlamydia pneumoniae.
XX
XX      MO9927105-A2.
XX      03-JUN-1999.
PD
XX      20-NOV-1998; 98WO-1B01890.
XX
XX      04-NOV-1998; 98US-0107078.
PR      21-NOV-1997; 97FR-0014673.
XX
XX      (GEST ) GENSET.
XX
XX      Griffais R;
XX      WPI: 1999-357842/30.
XX      Genome sequence of Chlamydia pneumoniae
PT
XX      Page 1737; Disclosure: 1912pp; English.
XX
XX      AAK91991-X97517 represent PCR primers used to amplify open reading
CC      frames and other nucleic acid sequences from the genome of
CC      Chlamydia pneumoniae (see AAK91990). C. pneumoniae causes respiratory
CC      disease such as pneumonia and bronchitis and is thought to be a
CC      contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC      otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC      by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC      AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC      containing C. pneumoniae nucleotide sequences can also be used as
CC      immunogenic compositions, especially where the vector directs the
CC      expression of a neutralising epitope of C. pneumoniae.
XX
XX      Sequence 20 BP; 7 A; 4 C; 5 G; 4 T; 0 other;
SQ
Query Match      1.2%; Score 20; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      72  CCTTTGGAATTCGCTC  91
      |||||||
DB      20  CCTTTGGAATTCGCTC  1

RESULT 16
AAK95297/C
ID      AAK95297 standard; DNA: 20 BP.
XX
XX      AAK95297;
XX
XX      13-SEP-1999 (first entry)
XX
XX      PCR primer used to amplify an ORF of Chlamydia pneumoniae.
DE
XX      Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW      sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW      vaccine; neutralising epitope; PCR primer; ss.
XX
XX      Synthetic.
OS      Chlamydia pneumoniae.
XX
XX      MO9927105-A2.
XX      03-JUN-1999.
XX

```


PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R:
 XX WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS Page 1737; Disclosure: 1912pp; English.

XX AAX91991-X97517 represent PCR primers used to amplify open reading
 CC frames and other nucleic acid sequences from the genome of
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
 CC disease such as pneumonia and bronchitis and is thought to be a
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
 CC containing C. pneumoniae nucleotide sequences can also be used as
 CC immunogenic compositions, especially where the vector directs the
 CC expression of a neutralising epitope of C. pneumoniae.

SQ Sequence 20 BP; 6 A; 7 C; 3 G; 4 T; 0 other;

Query Match 1.2%; Score 20; DB 20; Length 20;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1018 GTGAGTTCATGGGGAACCTTC 1037
 ||||||||||||||||
 Db 20 GTGAGTTCATGGGGAACCTTC 1

RESULT 17
 AAX95298/C
 ID AAX95298 standard; DNA; 20 BP.
 XX
 AC AAX95298;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; PCR primer; ss.
 XX
 PA (GEST) GENSET.
 OS Synthetic.
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PS 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R:
 XX WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS Page 1737; Disclosure: 1912pp; English.

XX AAX91991-X97517 represent PCR primers used to amplify open reading
 CC frames and other nucleic acid sequences from the genome of
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
 CC disease such as pneumonia and bronchitis and is thought to be a
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
 CC containing C. pneumoniae nucleotide sequences can also be used as
 CC immunogenic compositions, especially where the vector directs the
 CC expression of a neutralising epitope of C. pneumoniae.

SQ Sequence 20 BP; 9 A; 5 C; 5 G; 1 T; 0 other;

Query Match 1.2%; Score 20; DB 20; Length 20;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 ATTCGCTCTTCTTGTGGC 102
 ||||||||||||||||
 Db 20 ATTCGCTCTTCTTGTGGC 1

RESULT 18
 AAX92828
 ID AAX92828 standard; DNA; 20 BP.
 XX
 AC AAX92828;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; PCR primer; ss.
 XX
 PA (GEST) GENSET.
 OS Synthetic.
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PS 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 OS Synthetic.
 OS Chlamydia pneumoniae.
 XX
 PI Griffais R:
 XX WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS Page 1542; Disclosure: 1912pp; English.

XX AAX91991-X97517 represent PCR primers used to amplify open reading
 CC frames and other nucleic acid sequences from the genome of
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
 CC disease such as pneumonia and bronchitis and is thought to be a
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
 CC containing C. pneumoniae nucleotide sequences can also be used as
 CC immunogenic compositions, especially where the vector directs the
 CC expression of a neutralising epitope of C. pneumoniae.

SQ Sequence 20 BP; 5 A; 3 C; 7 G; 5 T; 0 other;

```

Query Match      1.2%; Score 20; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1244 GGTTGCGAGCTATACAGA 1263
DB 1 GGTTGCGAGCTATACAGA 20

RESULT 19
AAx92823
ID AAX92823 standard; DNA; 20 BP.
XX
AC AAX92823;
XX
DT 13-SEP-1999 (first entry)
XX
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; PCR primer; ss.
XX
OS Synthetic.
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078-
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.
XX
PI Grifflais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1541; Disclosure: 1912pp; English.
XX
CC AAX91991-x97517 represent PCR primers used to amplify open reading
CC frames and other nucleic acid sequences from the genome of
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
CC disease such as pneumonia and bronchitis and is thought to be a
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC containing C. pneumoniae nucleotides sequences can also be used as
CC immunogenic compositions, especially where the vector directs the
CC expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 20 BP; 4 A; 6 C; 4 G; 6 T; 0 other;

Query Match      1.2%; Score 20; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1446 GGAGCTATGACCCCTTATCT 1465
DB 1 GGAGCTATGACCCCTTATCT 20

RESULT 20
AAx92826
ID AAX92826 standard; DNA; 20 BP.
XX

```

```

AC AAX92826;
XX
DT 13-SEP-1999 (first entry)
XX
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; PCR primer; ss.
XX
OS Synthetic.
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078-
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.
XX
PI Grifflais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1542; Disclosure: 1912pp; English.
XX
CC AAX91991-x97517 represent PCR primers used to amplify open reading
CC frames and other nucleic acid sequences from the genome of
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
CC disease such as pneumonia and bronchitis and is thought to be a
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC containing C. pneumoniae nucleotides sequences can also be used as
CC immunogenic compositions, especially where the vector directs the
CC expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 20 BP; 2 A; 4 C; 5 G; 9 T; 0 other;

Query Match      1.2%; Score 20; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 924 CYCTTGATTATGCGCTATGG 943
DB 1 CYCTTGATTATGCGCTATGG 20

RESULT 21
AAF61373
ID AAF61373 standard; cDNA; 1770 BP.
XX
AC AAF61373;
XX
DT 25-MAY-2001 (first entry)
XX
DE A. thaliana chloroplast AtP/ADP translocator cDNA.
XX
KW ATP/ADP translocator; chloroplast; plant; adenosine triphosphate;
KW adenosine diphosphate; animal feed; energy; biosynthesis;
KW nutritional value; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200120009-A1.
XX

```

PD 22-MAR-2001.
XX
PF 05-AUG-2000; 2000MO-EP07625.
XX
PR 15-SEP-1999; 99DE-1044212.
XX
PA (BADI) BASF AG.
XX
PI Reindl A, Geigenberger PL, Neuhaus H, Graeve-Kampfenkel K;
PI Moehlmann T, Tjaden J;
XX
DR WPI: 2001-244814/25.
XX
PT Transformed plant with an altered content of essential amino acids and
PT having improved nutritional quality, for use as an animal feed,
PT comprises a modification in the ATP/ADP translocator gene
XX
XX
PS Claim 6; Fig 1; 32pp; German.
XX
CC This invention describes a novel transformed plant (A), and its
CC descendants, in which the regulatory sequences and/or the gene copy
CC number of an ATP/ADP (adenosine tri-/di-phosphate) translocator gene
CC (I) has been altered so that, compared with the wild-type plant, it
CC has a different content of one or more amino acids (aa). The invention
CC also describes (1) (I), for use in plants, having an approximately 1.7
CC kb sequence (S1), given in the specification, from Arabidopsis thaliana,
CC EMBL 249227; (2) a genetic construct (GC) containing the (I) and operably
CC linked regulatory sequences; (3) a vector containing the (I) or GC; and
CC (4) seeds, tissues, cells and replicative materials of (A). (A) are
CC useful for human or animal feeding, also (including their cells, tissues
CC and extracts) in agriculture, the animal feed and pharmaceutical
CC industries, and the health service. Altering the activity of (I)
CC increases the amount of ATP in chloroplasts and the amount of energy
CC available for biosynthesis. Specifically (A) have increased contents of
CC one or more essential aa, so are of greater nutritional value than
CC wild-type plants.
XX
SQ Sequence 1770 BP; 441 A; 383 C; 431 G; 515 T; 0 other;

Query Match 1.2%; Score 20; DB 22; Length 1770;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1308 AAAGAAATGGCCTATATCC 1327
Db 1489 AAAGAAATGGCCTATATCC 1508

RESULT 22
AAK77984/C
ID AAK77984 standard; DNA; 2287 BP.
XX
AC AAK77984;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32796.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-020515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227099.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0246174.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Disclosure: SEQ ID NO 32796; 3071pp + Sequence Listing; English.
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (1)

CC amino acid sequences given in AAM62170 to AAM91921. (1) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (1)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (1) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (1) by expressing inactive proteins or to
 CC supplement the patient's own production of (1). Additionally, (1)
 CC polynucleotides may be used to produce the secreted (1), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (1) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK67694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM62169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 2287 BP; 843 A; 386 C; 428 G; 630 T; 0 other;
 Query Match 1.2%; Score 20; DB 22; Length 2287;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 714 ATTCTTTACGCTTTTGAT 733
 Db 645 ATTCTTTACGCTTTTGAT 626
 ||||||||||||||||
 RESULT 23
 AAK24652
 ID AAK24652 standard; DNA; 33780 BP.
 XX
 AC AAK24652;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Nucleotide sequence of disrupted in schizophrenia 1 (DIS1) intron 8.
 XX
 KW Disrupted in schizophrenia 1 gene; DIS1 gene; schizophrenia;
 KW psychiatric disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200140301-A2.
 XX
 PD 07-JUN-2001.
 XX
 PE 28-NOV-2000; 2000WO-EP11915.
 XX
 PF 01-DEC-1999; 99EP-0309667.
 XX
 PA (ALKU) AKZO NOBEL NV.
 PA (MED1-) MEDICAL RES COUNCIL.
 PA (UYED-) UNIV EDINBURGH.
 XX
 PI Porteous D, Millar K, Blackwood D;
 XX
 DR WPI; 2001-374796/39.
 XX
 PT Novel isolated polynucleotide which surrounds a breakpoint on
 PT chromosome 1 involved in a balanced t(1;11) (q42.1;q14.3)
 PT translocation, and its encoded proteins, useful as medicament for
 PT treating psychiatric disorders -
 XX
 PS Disclosure: Page 39-50; 51pp; English.
 XX
 CC The present sequence represents intron 8 of a human disrupted in
 CC schizophrenia 1 (DIS1) gene. DIS1 is disrupted by a (1;11)(q42.1;q14.3)
 CC translocation which segregates with schizophrenia. The DIS1 gene consists
 CC of 13 exons. Exon 11 contains an alternative splice site which removed 66
 CC nucleotides from the open reading frame. The final intron of DIS1 belongs
 CC to the rare A/C class of introns. DIS1 polypeptide fragments and
 CC antibodies are useful for in vitro diagnosis of a psychiatric disorder.

CC DIS1 protein is useful in a screening assay for identifying new drugs for
 CC treating psychiatric disorders. DIS1 polypeptides and polynucleotides are
 CC useful for preventing and/or treating diseases related to schizophrenia.
 XX

XX Sequence 33780 BP; 9140 A; 6838 C; 7205 G; 10408 T; 189 other;

Query Match 1.2%; Score 20; DB 22; Length 33780;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 TTTTATTACTCTCTTGTT 932
 |||||||
 DB 21958 TTTTATTACTCTCTTGTT 21977

RESULT 24
 ABR83575
 ID ABR83575 standard; cDNA; 136284 BP.

XX ABR83575;

DT 14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #146.

XX Human: ss: granulocytic cell; DNA chip: bacterial infection;
 KM viral infection; parasitic infection; protozoal infection;
 KM fungal infection; sterile inflammatory disease; psoriasis;
 KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KM adult respiratory distress syndrome; inflammatory bowel disease;
 KM Crohn's disease; ulcerative colitis; periodontal disease;
 KM granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

PF 03-OCT-2001; 2001WO-US30821.

PR 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

DR WPI; 2002-435328/46.

XX
 PT Detecting granulocyte activation by detecting differential expression
 of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 drug toxicity

PS Claim 1: SEQ ID NO 146; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,

CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 136284 BP; 41693 A; 28668 C; 28000 G; 37923 T; 0 other;

Query Match 1.2%; Score 20; DB 24; Length 136284;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 187 ATGCAAGACACTCTTATT 206
 |||||||
 DB 64587 ATGCAAGACACTCTTATT 64606

RESULT 25

AAH12882
 ID AAH12882 standard; cDNA; 562 BP.

XX AAH12882;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:9717.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection
 XX and/or diagnosis of the abnormality of the proteins encoded by the
 XX full-length cDNAs -
 XX Claim 3: SEQ ID 9717; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
SQ Sequence 562 BP; 177 A; 113 C; 79 G; 182 T; 11 other;

Query Match 1.2%; Score 19; DB 22; Length 562;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 628 ATATTCTTACTAGCTTC 646
|||||
DB 381 ATATTCTTACTAGCTTC 399

RESULT 26
AAZ0940
ID AAZ0940 standard; DNA; 644 BP.
XX
AC AAZ0940;
XX
DT 05-MAY-1999 (first entry)
XX
DE Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX enzyme production; ds.
OS
XX Treponema pallidum.
XX
PN WO9859034-A2.
XX
PD 30-DEC-1998.
XX
PE 23-JUN-1998; 98WO-US13041.
XX
PR 24-JUN-1997; 97US-0050667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fraser CM;
XX
DR WPI: 1999-081273/07.
XX
XX New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX
PS Claim 1: Page 981-982; 1150pp; English.
XX
CC AAZ20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the

CC production of biosynthetic products such as enzymes.
XX
SQ Sequence 644 BP; 206 A; 129 C; 128 G; 178 T; 3 other;

Query Match 1.2%; Score 19; DB 20; Length 644;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 835 AAAAGGGGAAAAAGGTGC 853
|||||
DB 345 AAAAGGGGAAAAAGGTGC 363

RESULT 27
AAI96504/C
ID AAI96504 standard; CDNA; 764 BP.
XX
AC AAI96504;
XX
DT 13-NOV-2001 (first entry)
XX

DE Human neuroblastoma expressed polynucleotide SEQ ID NO 2579.

KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP01629.
XX
PR 07-MAR-2000; 2000JP-0159195.
XX
PA (CHTB-) CHIBA PREFECTURE.
XX
PI (HISM) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX

DR WPI: 2001-565584/63.
XX
XX
XX Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents -
XX
PS Claim 1: Page 1895; 2979pp; Japanese.
XX
XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
SQ Sequence 764 BP; 217 A; 131 C; 158 G; 227 T; 31 other;

Query Match 1.2%; Score 19; DB 22; Length 764;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 704 TCCTGGGGAATTTCTTTA 722
|||||
DB 422 TCCTGGGGAATTTCTTTA 404

RESULT 28
AAS74111/C
ID AAS74111 standard; CDNA; 790 BP.
XX
AC AAS74111;
XX

DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #9915.
 DE
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG09924.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1: SEQ ID No 9915; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 790 BP; 279 A; 132 C; 169 G; 210 T; 0 other;
 XX
 Query Match 1.2%; Score 19; DB 23; Length 790;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 900 TCTCTTATATCTTTTAT 918
 ||||||||||||||||
 Db 728 TCTCTTATATCTTTTAT 710

DE Haemophilus influenzae (Hird) Hrb gene upstream sequence, SEQ ID:64.
 XX
 KW Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
 KW genetically modified; protective antigen expression; LPS detoxification;
 KW LPS; Lipid A; homologous recombination vector; immunisation;
 KW immunoprotective; non-toxic; paediatric; ds.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO200109350-A2.
 XX
 PD 08-FEB-2001.
 XX
 PF 31-JUL-2000; 2000WO-EP07424.
 XX
 PR 03-AUG-1999; 99GB-0018319.
 XX
 PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;
 PI Poolman J, Thiry G, Thonard J, Voet P;
 XX
 DR WPI: 2001-138654/14.
 DR
 XX
 PT New isolated polynucleotide useful for outer membrane vesicle
 PT preparation from Gram-negative bacterial strain for vaccination of
 PT microbial infections -
 XX
 PS Claim 46: Page 93-94; 128bp; English.
 XX
 CC The invention relates to a genetically-engineered outer membrane vesicle
 CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
 CC The blebs of the invention are improved with respect to their
 CC immunogenicity and toxicity by the introduction of one or more genetic
 CC changes to the chromosome of the bacterium from which the blebs are
 CC derived. The changes made include the upregulation of protective antigen
 CC expression, the downregulation of immunodominant non-protective antigen
 CC expression, and genetic changes which result in detoxification of the
 CC lipid A moiety of lipopolysaccharide (LPS). The invention also
 CC encompasses modified Gram-negative bacterial strains from which the bleb
 CC preparations are made, a vector suitable for performing recombination
 CC events (for the generation of the modified bacterial strains),
 CC bacterially-derived nucleic acid sequences used in such a vector, and an
 CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
 CC cell vaccine suitable for paediatric use. The bleb preparation is useful
 CC in the manufacture of a medicament for immunising a human host against a
 CC disease caused by infection of one or more of the following: Neisseria
 CC meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
 CC catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
 CC pneumonia. The invention may also be used to provide immunisation against
 CC the influenza virus. Bacterially derived nucleotide sequences of the
 CC invention are used in the performance of homologous recombination events
 CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
 CC increase or decrease expression of that gene. Immunoprotective and
 CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
 CC are more immunogenic, less toxic and safer, and are particularly useful
 CC for paediatric use. The present sequence represents a specifically
 CC claimed Haemophilus influenzae nucleic acid sequence.
 XX
 SQ Sequence 1001 BP; 301 A; 225 C; 168 G; 307 T; 0 other;
 XX
 Query Match 1.2%; Score 19; DB 22; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 150 TTCTCTGTATTACATTTA 168
 ||||||||||||||||
 Db 275 TTCTCTGTATTACATTTA 293

RESULT 30
 AAK37818
 ID AAK37818 standard; DNA; 1001 BP.
 XX
 DT 04-MAY-2001 (first entry)
 XX

XX	ABK37818;	
AC		
XX		
DT	08-MAY-2002 (first entry)	
XX		
DE	DNA sequence upstream of htrb #3 gene.	
XX		
XX	Upstream sequence: ds; Antibacterial; vaccine; bleb;	
KM	Gram-negative bacteria; outer membrane; LPS; lipopolysaccharide;	
KW	meningitis; bacteraemia; otitis media; pneumonia; chronic bronchitis;	
XX	sinusitis.	
OS	Haemophilus influenzae H1rD.	
XX		
PN	WO200209746-A2.	
XX		
PD	07-FEB-2002.	
XX		
PF	31-JUL-2001; 2001WO-EP08857.	
XX		
PR	31-JUL-2000; 2000EP-0956369.	
PR	08-FEB-2001; 2001GB-0003170.	
XX		
PA	(SMK) SMTKHLNE BEECHAM BIOLOGICALS.	
XX		
PI	Berthet FJ, Dalemans W, Denoel P, Dequesne G, Feron C, Garcon N;	
PI	Lobet Y, Poolman J, Thiry G, Thonnard J, Voet P;	
XX		
XX	WPI, 2002-188688/24.	
DR		
XX		
XX	New immunogenic composition comprising an antigen derived from a	
PT	pathogen and a bleb preparation from Neisseria meningitidis, useful as	
PT	a vaccine for treating or preventing disease caused by the pathogen -	
XX		
PS	Disclosure: Page 96-97; 125pp; English.	
XX		
XX	The invention relates to an immunogenic composition comprising an antigen	
CC	derived from a pathogen capable of protecting a host against the	
CC	pathogen, mixed with an adjuvant comprising a bleb preparation derived	
CC	from a Gram-negative bacterial strain. The immunogenic composition	
CC	consists of N. meningitidis B blebs or N. meningitidis C polysaccharide	
CC	antigen. The blebs (derived from the outer membrane) may also	
CC	have their toxic lipopolysaccharide (LPS) content reduced using	
CC	heterologous down regulating sequences for LPS pathway genes or by	
CC	up regulating genes involved in LPS synthesis suppression, by	
CC	a promoter replacement technique. The immunogenic preparation is useful	
CC	in the manufacture of a medicament for the treatment of a disease caused	
CC	by the pathogen from which the antigen is derived (e.g. from	
CC	Neisseria, meningitis and bacteraemia, from Moraxella, otitis media and	
CC	pneumonia, and from H. influenzae chronic bronchitis, sinusitis,	
CC	pneumonia and otitis media). The bleb derived from M. catarrhalis or	
CC	from a non-typeable H. influenzae is useful as an adjuvant in an	
CC	immunogenic composition comprising one or more pneumococcal capsular	
CC	polysaccharides or protein antigens. The present sequence is an	
CC	upstream sequence from an N. meningitidis, H. influenzae or M. catarrhalis	
CC	gene involved in LPS biosynthesis, which either up regulates or down	
CC	regulates sequences to which it is attached.	
XX		
XX		
SQ	Sequence 1001 BP: 301 A: 225 C: 168 G: 307 T: 0 other:	
	Query Match 1.2%; Score 19; DB 24; Length 1001;	
	Best Local Similarity 100.0%; Pred. No. 53;	
	Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	150 TTCTCTGTATTTACATTTA 168	
DB	275 TTCTCTGTATTTACATTTA 293	

RESULT 31

ABV25656/C

ID ABV25656 standard; cDNA; 1362 BP.

ABK37818;

08-MAY-2002 (first entry)

DNA sequence upstream of htrb #3 gene.

Upstream sequence: ds; Antibacterial; vaccine; bleb;

Gram-negative bacteria; outer membrane; LPS; lipopolysaccharide;

meningitis; bacteraemia; otitis media; pneumonia; chronic bronchitis;

sinusitis.

Haemophilus influenzae H1rD.

WO200209746-A2.

07-FEB-2002.

31-JUL-2001; 2001WO-EP08857.

31-JUL-2000; 2000EP-0956369.

08-FEB-2001; 2001GB-0003170.

(SMK) SMTKHLNE BEECHAM BIOLOGICALS.

Berthet FJ, Dalemans W, Denoel P, Dequesne G, Feron C, Garcon N;

Lobet Y, Poolman J, Thiry G, Thonnard J, Voet P;

WPI, 2002-188688/24.

New immunogenic composition comprising an antigen derived from a

pathogen and a bleb preparation from Neisseria meningitidis, useful as

a vaccine for treating or preventing disease caused by the pathogen -

Disclosure: Page 96-97; 125pp; English.

The invention relates to an immunogenic composition comprising an antigen

derived from a pathogen capable of protecting a host against the

pathogen, mixed with an adjuvant comprising a bleb preparation derived

from a Gram-negative bacterial strain. The immunogenic composition

consists of N. meningitidis B blebs or N. meningitidis C polysaccharide

antigen. The blebs (derived from the outer membrane) may also

have their toxic lipopolysaccharide (LPS) content reduced using

heterologous down regulating sequences for LPS pathway genes or by

up regulating genes involved in LPS synthesis suppression, by

a promoter replacement technique. The immunogenic preparation is useful

in the manufacture of a medicament for the treatment of a disease caused

by the pathogen from which the antigen is derived (e.g. from

Neisseria, meningitis and bacteraemia, from Moraxella, otitis media and

pneumonia, and from H. influenzae chronic bronchitis, sinusitis,

pneumonia and otitis media). The bleb derived from M. catarrhalis or

from a non-typeable H. influenzae is useful as an adjuvant in an

immunogenic composition comprising one or more pneumococcal capsular

polysaccharides or protein antigens. The present sequence is an

upstream sequence from an N. meningitidis, H. influenzae or M. catarrhalis

gene involved in LPS biosynthesis, which either up regulates or down

regulates sequences to which it is attached.

Sequence 1001 BP: 301 A: 225 C: 168 G: 307 T: 0 other:

Query Match 1.2%; Score 19; DB 24; Length 1001;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

150 TTCTCTGTATTTACATTTA 168

|||||

275 TTCTCTGTATTTACATTTA 293

```

AC ABV25656;
AD
AE
AF
AG
AH
AI
AJ
AK
AL
AM
AN
AO
AP
AQ
AR
AS
AT
AU
AV
AW
AX
AY
AZ
BA
BB
BC
BD
BE
BF
BG
BH
BI
BJ
BK
BL
BM
BN
BO
BP
BQ
BR
BS
BT
BU
BV
BW
BX
BY
BZ
CA
CB
CC
CD
CE
CF
CG
CH
CI
CJ
CK
CL
CM
CN
CO
CP
CQ
CR
CS
CT
CU
CV
CW
CX
CY
CZ
DA
DB
DC
DD
DE
DF
DG
DH
DI
DJ
DK
DL
DM
DN
DO
DP
DQ
DR
DS
DT
DU
DV
DW
DX
DY
EZ
FA
FB
FC
FD
FE
FF
FG
FH
FI
FJ
FK
FL
FM
FN
FO
FP
FQ
FR
FS
FT
FU
FV
FW
FX
FY
FZ
GA
GB
GC
GD
GE
GF
GG
GH
GI
GJ
GK
GL
GM
GN
GO
GP
GQ
GR
GS
GT
GU
GV
GW
GX
GY
GZ
HA
HB
HC
HD
HE
HF
HG
HH
HI
HJ
HK
HL
HM
HN
HO
HP
HQ
HR
HS
HT
HU
HV
HW
HX
HY
HZ
IA
IB
IC
ID
IE
IF
IG
IH
II
IJ
IK
IL
IM
IN
IO
IP
IQ
IR
IS
IT
IU
IV
IW
IX
IY
IZ
JA
JB
JC
JD
JE
JF
JG
JH
JI
JJ
JK
JL
JM
JN
JO
JP
JQ
JR
JS
JT
JU
JV
JW
JX
JY
JZ
KA
KB
KC
KD
KE
KF
KG
KH
KI
KJ
KK
KL
KM
KN
KO
KP
KQ
KR
KS
KT
KU
KV
KW
KX
KY
KZ
LA
LB
LC
LD
LE
LF
LG
LH
LI
LJ
LK
LL
LM
LN
LO
LP
LQ
LR
LS
LT
LU
LV
LW
LX
LY
LZ
MA
MB
MC
MD
ME
MF
MG
MH
MI
MJ
MK
ML
MM
MN
MO
MP
MQ
MR
MS
MT
MU
MV
MW
MX
MY
MZ
NA
NB
NC
ND
NE
NF
NG
NH
NI
NJ
NK
NL
NM
NO
NP
NQ
NR
NS
NT
NU
NV
NW
NX
NY
NZ
OA
OB
OC
OD
OE
OF
OG
OH
OI
OJ
OK
OL
OM
ON
OO
OP
OQ
OR
OS
OT
OU
OV
OW
OX
OY
OZ
PA
PB
PC
PD
PE
PF
PG
PH
PI
PJ
PK
PL
PM
PN
PO
PP
PQ
PR
PS
PT
PU
PV
PW
PX
PY
PZ
QA
QB
QC
QD
QE
QF
QG
QH
QI
QJ
QK
QL
QM
QN
QO
QP
QQ
QR
QS
QT
QU
QV
QW
QX
QY
QZ
RA
RB
RC
RD
RE
RF
RG
RH
RI
RJ
RK
RL
RM
RN
RO
RP
RQ
RR
RS
RT
RU
RV
RW
RX
RY
RZ
SA
SB
SC
SD
SE
SF
SG
SH
SI
SJ
SK
SL
SM
SN
SO
SP
SQ
SR
SS
ST
SU
SV
SW
SX
SY
SZ
TA
TB
TC
TD
TE
TF
TG
TH
TI
TJ
TK
TL
TM
TN
TO
TP
TQ
TR
TS
TT
TU
TV
TW
TX
TY
TZ
UA
UB
UC
UD
UE
UF
UG
UH
UI
UJ
UK
UL
UM
UN
UO
UP
UQ
UR
US
UT
UU
UV
UW
UX
UY
UZ
VA
VB
VC
VD
VE
VF
VG
VH
VI
VJ
VK
VL
VM
VN
VO
VP
VQ
VR
VS
VT
VU
VV
VW
VX
VY
VZ
WA
WB
WC
WD
WE
WF
WG
WH
WI
WJ
WK
WL
WM
WN
WO
WP
WQ
WR
WS
WT
WU
WV
WW
WX
WY
WZ
XA
XB
XC
XD
XE
XF
XG
XH
XI
XJ
XK
XL
XM
XN
XO
XP
XQ
XR
XS
XT
XU
XV
XW
XX
XY
XZ
YA
YB
YC
YD
YE
YF
YG
YH
YI
YJ
YK
YL
YM
YN
YO
YP
YQ
YR
YS
YT
YU
YV
YW
YX
YY
YZ
ZA
ZB
ZC
ZD
ZE
ZF
ZG
ZH
ZI
ZJ
ZK
ZL
ZM
ZN
ZO
ZP
ZQ
ZR
ZS
ZT
ZU
ZV
ZW
ZX
ZY
ZZ

```


KW	Human; microglobulin transcriptional control factor 30;
KW	recombinant production; malignant tumour; cancer; blood disease;
KW	HIV infection; human immunodeficiency virus; immune disorder;
KW	inflammatory condition; cytosolic; anti-HIV; anti-inflammatory;
KM	immunomodulator; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	99..908
FT	CDS
FT	/tag= a
FT	/product= "Human microglobulin transcriptional control
XX	factor 30"
PN	MO200188151-A1.
XX	
PD	22-NOV-2001.
XX	
XX	08-MAY-2001; 2001WO-CN00709.
PR	09-MAY-2000; 2000CN-0115610.
XX	
PA	(SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX	
PI	Mao Y, Xie Y;
XX	
DR	WPI; 2002-034729/04.
DR	P-PSDB; AAM52672.
XX	
PT	Human Microglobulin Transcriptional Control Factor 30 and encoding
PT	polynucleotide, used in diagnosis and treatment of malignant tumors,
PT	hemopathy, human immunodeficiency virus infection, immunological
PT	diseases and inflammation -
XX	
PS	Claim 6; Page 30-31; 38pp; Chinese.
XX	
CC	This sequence represents cDNA encoding human microglobulin
CC	transcriptional control factor 30. The protein has a molecular weight of
CC	30 kD. The invention relates to human microglobulin transcriptional
CC	control factor 30 (AAM52672), nucleic acids encoding it (ABA02319), and
CC	a method for the recombinant production of microglobulin transcriptional
CC	control factor 30. The present invention additionally discloses an
CC	antagonist of microglobulin transcriptional control factor 30 for
CC	therapeutic use, and an antibody which specifically binds to
CC	microglobulin transcriptional control factor 30. Microglobulin
CC	transcriptional control factor 30, and nucleotides which encode it may
CC	be used for treating a variety of diseases, such as malignant tumours,
CC	blood diseases, HIV (human immunodeficiency virus) infection, immune
CC	disorders and inflammatory conditions. The protein may also be used to
CC	screen for modulators of its activity or for peptide fingerprinting
CC	identification. The polynucleotide can be used as a primer for nucleic
CC	acid amplification reactions or as a probe for hybridisation reactions,
CC	or in producing gene chips or microarrays.
XX	
SQ	Sequence 1604 BP; 572 A; 230 C; 269 G; 533 T; 0 other:
	Query Match 1.2%; Score 19; DB 24; Length 1604;
	Best Local Similarity 100.0%; Pred. No. 53;
	Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	900 TCCTCCTTATTCCTTTAT 918
Db	621 TTCCTTATTCCTTTAT 603
RESULT 33	
ABLA1351	
ID	ABLA1351 standard; cDNA; 1689 BP.
AC	ABLA1351;
XX	
XT	22-MAY-2002. (first entry)
XX	

DE	Human glycosyl-phosphatidyl inositol glycan F10.89 encoding cDNA.	
XX		
KW	Human: glycosyl-phosphatidyl inositol glycan F10.89;	
KW	embryonic development malformation; immunological dysfunction;	
KW	hormonal metabolic disturbance; gene; ss.	
OS		
XX		
XX	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	945..1244
FT		/*tag= a
FT		/product= "glycosyl-phosphatidyl inositol glycan F10.89"
XX		
PN	CN1326960-A.	
XX		
PD	19-DEC-2001.	
XX		
PF	05-JUN-2000; 2000CN-0116326.	
XX		
PR	05-JUN-2000; 2000CN-0116326.	
XX		
PA	(BODE-) BODE GENE DEV CO LTD SHANGHAI.	
XX		
PI	Mao Y, Xie Y;	
XX		
DR	WPI: 2002-206969/27.	
XX		
DR	P-PSDB: ABB04400.	
XX		
PT	New polypeptide-glycosyl-phosphatidyl inositol glycan F10.89 and	
XX	polynucleotide encoding the polypeptide -	
PS	Claim 6; Page 27-28 (Disclosure): 35pp; Chinese.	
XX		
CC	The invention relates to human glycosyl-phosphatidyl inositol glycan	
CC	F10.89, the polynucleotide encoding this polypeptide and DNA recombinant	
CC	processes to produce the polypeptide. The present invention also	
CC	discloses the method of applying the polypeptide in treating various	
CC	diseases, such as embryonic development malformation, immunological	
CC	dysfunction disease and hormonal metabolic disturbance disease. The	
CC	present invention also discloses the antagonist for resisting the	
CC	polypeptide and its treatment effect. The present invention also	
CC	discloses the application of the polynucleotide for encoding	
XX	glycosyl-phosphatidyl inositol glycan F10.89.	
XX		
SQ	Sequence 1689 BP; 541 A; 345 C; 348 G; 455 T; 0 other;	
XX		
Query Match	1.2%; Score 19; DB 24; Length 1689;	
Best Local Similarity	100.0%; Pred. NO. 53;	
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	187 ATACAAAGACACTCTTAT 205	
DB	311 ATACAAAGACACTCTTAT 329	
XX		
RESULT 34		
ID	AAH15049/C	
XX	AAH15049 standard; cDNA; 1833 BP.	
XX		
AC	AAH15049;	
XX		
DT	26-JUN-2001 (first entry)	
XX		
DE	Human cDNA sequence SEQ ID NO:13039.	
XX		
KW	Human: primer; detection; diagnosis; antisense therapy; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1074617-A2.	
XX		
XX	07-FEB-2001.	
XX		

PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length CDNA defined in the specification, and/or the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length CDNA -
 XX
 PS Claim 8; SEQ ID 13039; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length CDNA defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length CDNA. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length CDNA. The primers allow obtaining of the full-length
 CC CDNA easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1833 BP; 595 A; 309 C; 313 G; 616 T; 0 other;
 XX
 Query Match 1.2%; Score 19; DB 22; Length 1833;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 628 ATATTCTTCTACTAGCTTC 646
 ||||||||||||||||
 DB 1453 ATATTCTTCTACTAGCTTC 1435
 RESULT 35
 ABL12876/C
 ID ABL12876 standard; CDNA: 2611 BP.
 XX
 AC ABL12876;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33110.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 DR

PD 27-SEP-2001.
 XX
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI: 2001-656860/75.
 DR P-PSDB: ABB68773.
 XX
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 33110; 21pp + Sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2611 BP; 855 A; 546 C; 540 G; 670 T; 0 other;
 XX
 Query Match 1.2%; Score 19; DB 23; Length 2611;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 584 CCACGACGACCAAGCGTTTC 602
 ||||||||||||||||
 DB 891 CCACGACGACCAAGCGTTTC 873
 RESULT 36
 ABL04386/C
 ID ABL04386 standard; CDNA: 3052 BP.
 XX
 AC ABL04386;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 7640.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB: ABB60283.
 DR

XX	PS	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -	Claim 1; SEQ ID NO 7640; 21np + Sequence Listing; English.
XX	CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB161716-AB130511), expressed DNA sequences (AB101840-AB161715) and the encoded proteins (ABB57737-ABB72072).	
XX	CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX	SO	Sequence 3052 BP; 792 A; 651 C; 705 G; 904 T; 0 other;	
XX	Query Match	1.2%; Score 19; DB 23; Length 3052;	
XX	Best Local Similarity	100.0%; Pred. No. 52;	
XX	Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	299 AAGCTAAGTATATTTTA 317		
DB	591 AAGCTAAGTATATTTTA 573		
XX	ABK13001/c		
XX	ABK13001 standard; cDNA; 3752 BP.		
XX	ABK13001;		
XX	23-APR-2002 (first entry)		
XX	DNA encoding SCAN zinc-finger protein subfamily protein 57.		
XX	SCAN zinc-finger protein subfamily protein 57; cytosolic; neuroprotectant; nootropic; cerebroprotective; tumour; nervous system disorder; neurodegenerative disease; neural disease; developmental disturbance; psychogenic disease; gene; ss.		
XX	Unidentified.		
XX	Key	Location/Qualifiers	
XX	CDS	131..1687	
XX	FT	/*tag= a	
XX	FT	/product= "SCAN zinc-finger protein subfamily protein 57"	
XX	MO200194530-A2.		
XX	13-DEC-2001.		
XX	14-MAY-2001; 2001WO-CN00774.		
XX	16-MAY-2000; 2000CN-0115706.		
XX	(SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.		
XX	Mao Y, Xie Y;		
XX	WI: 2002-083223/11.		
XX	P-PSDB; AAU11020.		
XX	SCAN zinc-finger subfamily protein 57 and encoding polynucleotide, for diagnosis and treatment of tumours of the nervous system, neurodegenerative disease, developmental disturbances, psychogenic disease and other neural diseases -		
XX	Claim 6; Page 30-32; 35pp; Chinese.		

XX	The invention relates to an isolated polypeptide (I) of SCAN zinc-finger
CC	protein subfamily protein 57 and the polynucleotide (I) encoding (II).
CC	(I) and (II) are used in diagnosis and treatment of tumours of the
CC	nervous system, neurodegenerative disease, developmental disturbances
CC	of the nervous system, psychogenic disease and other neural diseases.
CC	The present sequence represents the coding sequence of SCAN zinc-
CC	finger protein subfamily protein 57.
XX	
SQ	Sequence 3752 BP; 1205 A; 704 C; 754 G; 1089 T; 0 other;
Query Match	1.2%; Score 19; DB 24; Length 3752;
Best Local Similarity	100.0%; Pred. No. 52;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	628 ATATTTCTTACTAGCTTC 646
Dd	3376 ATATTCTTTACTAGCTTC 3358
RESULT 38	
ID	ABL24540/c
XX	ABL24540 standard; DNA; 4007 BP.
AC	
XX	ABL24540;
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 25093.
XX	
MW	Drosophila: developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ds.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
PS	
PS	Claim 1: SEQ ID NO 25093; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 4007 BP; 1201 A; 929 C; 887 G; 990 T; 0 other;
Query Match	1.2%; Score 19; DB 23; Length 4007;
Best local Similarity	100.0%; Pred. No. 52;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

QY 584 CCACGAGCAAGCGTTTC 602
 |||||||
 DB 3749 CCACGAGCAAGCGTTTC 3731

RESULT 39
 ABL11410/c
 ID ABL11410 standard; cDNA; 5586 BP.
 XX
 AC ABL11410;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 28712.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB; ABB67307.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 28712; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 5586 BP; 1581 A; 1196 C; 1164 G; 1645 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 5586;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 CCTTCATCAGTTTGGC 258
 |||||||
 DB 913 CCTTCATCAGTTTGGC 895

RESULT 40
 ABL03008/c
 ID ABL03008 standard; cDNA; 8034 BP.
 XX
 AC ABL03008;
 XX
 DT 26-MAR-2002 (first entry)

XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 3506.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB; ABB58905.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 3506; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 8034 BP; 2137 A; 1815 C; 1882 G; 2200 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 8034;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 299 AAGCTAGTATATTTTA 317
 |||||||
 DB 6827 AAGCTAGTATATTTTA 6809

Search completed: January 21, 2003, 21:03:30
 Job time : 2525 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 21:03:19 : Search time 59 seconds
(without alignments)
8508.981 Million cell updates/sec

Title: US-09-869-433-1

Perfect score: 1637
Sequence: 1 gaataaaactatcagaa.....gttgatccctactcgtcttt 1637

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	21	1.3	8220	2	US-08-568-459A-11
2	21	1.3	8220	2	US-08-487-826B-11
3	21	1.3	8220	4	US-09-210-288-11
4	21	1.3	19124	2	US-08-487-826B-13
5	18	1.1	3980	1	US-08-233-008A-1
6	18	1.1	3980	1	US-08-233-008A-5
7	18	1.1	6519	4	US-09-221-017B-311
8	18	1.1	6519	4	US-08-233-008A-7
9	18	1.1	14736	4	US-08-961-527-171
10	17	1.0	378	4	US-09-134-001C-2567
11	17	1.0	383	4	US-09-370-838-285
12	17	1.0	882	4	US-09-227-357-133
13	17	1.0	939	1	US-08-592-411-14
14	17	1.0	939	1	US-08-591-401-15
15	17	1.0	942	1	US-08-592-411-16
16	17	1.0	942	1	US-08-591-501-14
17	17	1.0	1316	2	US-08-871-924A-1
18	17	1.0	1316	4	US-09-361-900-1
19	17	1.0	1374	1	US-08-093-372-3
20	17	1.0	1380	2	US-08-748-947A-1
21	17	1.0	1614	4	US-09-110-959A-3
22	17	1.0	2561	4	US-09-270-542-101
23	17	1.0	2561	4	US-09-270-542-119
24	17	1.0	2880	1	US-08-158-189-1
25	17	1.0	3099	4	US-09-360-186-2
26	17	1.0	3810	2	US-08-475-844-8
27	17	1.0	3810	5	PCT-US95-08429-8

ALIGNMENTS

```
RESULT 1
US-08-568-459A-11
; Sequence 11, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-11

Query Match 1.3%; Score 21; DB 2; Length 8220;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AAGAAACCTTTGGAAAT 84
|
Db 5046 AAGAAACCTTTGGAAAT 5066

RESULT 2
US-08-487-826B-11
; Sequence 11, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
```

```
APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-11

Query Match 1.3%; Score 21; DB 2; Length 8220;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AAGAAACCTTTGGAAAT 84
|
Db 5046 AAGAAACCTTTGGAAAT 5066

RESULT 3
US-09-210-288-11
; Sequence 11, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210.288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 8220 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-11

Query Match          1.3%; Score 21; DB 4; Length 8220;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAACCTTTGGAAAT 84
|||||
Db 5046 AAGAAAACCTTTGGAAAT 5066

RESULT 4
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Shim, Kim I.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487.826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
```

```
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match          1.3%; Score 21; DB 2; Length 19124;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAACCTTTGGAAAT 84
|||||
Db 12366 AAGAAAACCTTTGGAAAT 12386

RESULT 5
US-08-233-008A-1
Sequence 1, Application US/08233008A
Patent No. 5578480
GENERAL INFORMATION:
APPLICANT: Khandke, Kiran M.
TITLE OF INVENTION: Methods For The Isolation And
Purification Of The Recombinantly Expressed Chondroitinase
TITLE OF INVENTION: I and II Enzymes From P. vulgaris
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470-8426
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.008A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 31,885-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3980 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 119..3181
US-08-233-008A-1

Query Match          1.1%; Score 18; DB 1; Length 3980;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 AAAGACACTCTTATGT 208
|||||
Db 2965 AAAGACACTCTTATGT 2982
```

```
RESULT 6
US-08-233-008A-5
; Sequence 5, Application US/08233008A
; Patent No. 5578480
; GENERAL INFORMATION:
; APPLICANT: Khandke, Kiran M.
; TITLE OF INVENTION: Methods For The Isolation And
; TITLE OF INVENTION: Purification Of The Recombinantly Expressed chondroitinase
; TITLE OF INVENTION: I and II Enzymes From P. vulgaris
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,008A
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,885-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3244
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 188..3181
; US-08-233-008A-5

Query Match 1.1%; Score 18; DB 1; Length 3980;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 AAAAGACACTTATTGT 208
Db 2965 AAAAGACACTTATTGT 2982

RESULT 7
US-09-221-017B-311
; Sequence 311, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/A098/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 311:
SEQUENCE CHARACTERISTICS:
LENGTH: 6321 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...6321
US-09-221-017B-311

Query Match 1.1%; Score 18; DB 4; Length 6321;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1163 TTCTTCGCTCTTGTAT 1180
Db 6152 TTCTTCGCTCTTGTAT 6169

RESULT 8
US-08-233-008A-7
; Sequence 7, Application US/08233008A
; Patent No. 5578480
; GENERAL INFORMATION:
; APPLICANT: Khandke, Kiran M.
; TITLE OF INVENTION: Methods For The Isolation And
; TITLE OF INVENTION: Purification Of The Recombinantly Expressed chondroitinase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```



```
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,008A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 31,885-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6519 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 3238..6276
US-08-233-008A-7

Query Match          1.1%; Score 18; DB 1: Length 6519;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 AAAAGACACTTATTGT 208
    |||||
DB 2965 AAAAGACACTTATTGT 2982

RESULT 9
US-08-961-527-171
Sequence 171, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 171:
SEQUENCE CHARACTERISTICS:
LENGTH: 14736 base pairs
TYPE: nucleic acid
```

```
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-171

Query Match          1.1%; Score 18; DB 4: Length 14736;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 ATCTACAGAAATTTGCTG 429
    |||||
DB 5312 ATCTACAGAAATTTGCTG 5329

RESULT 10
US-09-134-001C-2567
Sequence 2567, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2567
LENGTH: 378
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2567

Query Match          1.0%; Score 17; DB 4: Length 378;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1549 TTAAGACACAGAGTG 1565
    |||||
DB 29 TTAAGACACAGAGTG 45

RESULT 11
US-09-370-838-285/c
Sequence 285, Application US/09370838
Patent No. 6444425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Roadoh
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370,838
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/285,323
EARLIER FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 285
LENGTH: 383
TYPE: DNA
ORGANISM: Homo sapiens
US-09-370-838-285

Query Match          1.0%; Score 17; DB 4: Length 383;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 571 AAATTACAAAAATCCAC 587
```

Wed Jan 22 10:20:54 2003

Db 367 AATTACAAAATCCAC 351

```

|||||
RESULT 12
US-09-227-357-133/c
; Sequence 133, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010p1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785

```

```

; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO: 133
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (881)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-133

```

Query Match 1.0%; Score 17; DB 4; Length 882;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 ACCTAAGAAAGTCTG 134
 Db 875 ACCTAAGAAAGTCTG 859

```

RESULT 13
US-08-592-411-14/c
; Sequence 14, Application US/08592411
; Patent No. 5726032
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for the Efficient Production of
; TITLE OF INVENTION: 7-ADCA via 2-(Carboxymethylthio)scetyl-7-ADCA and
; NUMBER OF SEQUENCES: 17
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,411
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 939 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: No. 5726032ardia lactandurans
; INDIVIDUAL ISOLATE: ATCC 27382
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..939
; OTHER INFORMATION: /gene="cefE"
US-08-592-411-14

```

Query Match 1.0%; Score 17; DB 1; Length 939;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1366 TTGATGATGTCGCCGC 1382
 Db 914 TTGATGATGTCGCCGC 898

RESULT 14
US-08-591-501-15/C
Sequence 15, Application US/08591501
Patent No. 5795733
GENERAL INFORMATION:
APPLICANT: BOVENBERG, ROELOF ARY LANS
APPLICANT: KOEKMAN, BERTUS PIETER
APPLICANT: HOEKEMA, ANDREAS
APPLICANT: VAN DER LAAN, JAN METSKE
APPLICANT: VERWEIJ, JAN
APPLICANT: DE VROOM, ERIK
TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROPIONYL-7-ADCA
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/591,501
FILING DATE: 13-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: 24615-20065.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNPOBRSMSH
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 939 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-591-501-15
Query Match 1.0%; Score 17; DB 1; Length 939;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1366 TTGATGTAGTGGCCGC 1382
|||||
DB 914 TTGATGTAGTGGCCGC 898
RESULT 15
US-08-592-411-16/C
Sequence 16, Application US/08592411
Patent No. 5726032
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Process for the efficient production of
TITLE OF INVENTION: 7-ADCA via 2-(Carboxyethylthio)acetyl-7-ADCA and
TITLE OF INVENTION: 3-(Carboxymethylthio)propionyl-7-ADCA
NUMBER OF SEQUENCES: 17
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,411

FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 942 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: No. 5726032ardia lactamdurans
INDIVIDUAL ISOLATE: LC 411
FEATURE:
NAME/KEY: CDS
LOCATION: 1..942
OTHER INFORMATION: /gene="cefE"
OTHER INFORMATION: /citation= (11)
PUBLICATION INFORMATION:
AUTHORS: Coque, J J
AUTHORS: Martin, J F
AUTHORS: Liras, P
TITLE: Characterization and expression in
TITLE: Streptomyces lividans of cefD and cefE genes from
TITLE: No. 5726032ardia lactamdurans: the organization of the
TITLE: cephamycin gene cluster differs from that in
TITLE: Streptomyces clavuligerus
JOURNAL: Mol. Gen. Genet.
VOLUME: 236
PAGES: 453-458
DATE: 1993
US-08-592-411-16
Query Match 1.0%; Score 17; DB 1; Length 942;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1366 TTGATGTAGTGGCCGC 1382
|||||
DB 917 TTGATGTAGTGGCCGC 901
RESULT 16
US-08-591-501-14/C
Sequence 14, Application US/08591501
Patent No. 5795733
GENERAL INFORMATION:
APPLICANT: BOVENBERG, ROELOF ARY LANS
APPLICANT: KOEKMAN, BERTUS PIETER
APPLICANT: HOEKEMA, ANDREAS
APPLICANT: VAN DER LAAN, JAN METSKE
APPLICANT: VERWEIJ, JAN
APPLICANT: DE VROOM, ERIK
TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROPIONYL-7-ADCA
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/591,501
FILING DATE: 13-MAY-1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: 24615-20065.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSMWSH
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 942 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PUBLICATION INFORMATION:
AUTHORS: COQUE et al.,
JOURNAL: Mol. Gen. Genet.
VOLUME: 236
PAGES: 453-458
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 14: FROM 1 TO 942
US-08-591-501-14

Query Match 1.0%; Score 17; DB 1; Length 942;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1366 TTGATGTAGTGGCCGC 1382
DB 917 TTGATGTAGTGGCCGC 901

RESULT 17
US-08-871-924A-1/c
Sequence 1, Application US/08871924A
Patent No. 5932469
GENERAL INFORMATION:
APPLICANT: Hillman, Jeffrey D
TITLE OF INVENTION: No. 5932469e1 Antimicrobial Polypeptides and
TITLE OF INVENTION: Methods of Use
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney St
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II/DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,924A
FILING DATE: 06/10/97
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 920308.90649
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1316 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: -35_signal

LOCATION: 738..742
FEATURE:
NAME/KEY: -10_signal
LOCATION: 757..763
FEATURE:
NAME/KEY: RBS
LOCATION: 784..791
FEATURE:
NAME/KEY: CDS
LOCATION: 796..987
US-08-871-924A-1

Query Match 1.0%; Score 17; DB 2; Length 1316;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 863 AATGAATATGAAGATA 879
DB 260 AATGAATATGAAGATA 244

RESULT 18
US-09-361-900-1/c
Sequence 1, Application US/09361900
Patent No. 6391285
GENERAL INFORMATION:
APPLICANT: Hillman, Jeffrey D
TITLE OF INVENTION: No. 6391285e1 Antimicrobial Polypeptides and
TITLE OF INVENTION: Methods of Use
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney St
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II/DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/361,900
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/871,924
FILING DATE: 06/10/97
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 920308.90649
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1316 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: -35_signal
LOCATION: 738..742
FEATURE:
NAME/KEY: -10_signal
LOCATION: 757..763
FEATURE:
NAME/KEY: RBS
LOCATION: 784..791
FEATURE:

NAME/KEY: CDS
LOCATION: 796..987
US-09-361-900-1

Query Match 1.0%; Score 17; DB 4; Length 1316;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 863 AATGATATGAAAGATA 879
|||||
DB 260 AATGATATGAAAGATA 244

RESULT 19
US-08-093-372-3/C

Sequence 3, Application US/08093372
Patent No. 5530187
GENERAL INFORMATION:
APPLICANT: Lamb, Christopher J.
APPLICANT: Zhu, Qun
APPLICANT: Maher, Eileen A.
APPLICANT: Dixon, Richard A.
TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING MULTIPLE
TITLE OF INVENTION: DISEASE RESISTANCE GENES
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/093.372
FILING DATE: 16-JUL-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31.192
REFERENCE/DOCKET NUMBER: PA1 9391
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1374 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 52..1158
OTHER INFORMATION: /product="ALFALFA BETA-GLUCANASE"
US-08-093-372-3

Query Match 1.0%; Score 17; DB 1; Length 1374;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 861 AAATGAATATGAAAGA 877
|||||
DB 28 AAATGAATATGAAAGA 12

RESULT 20
US-08-748-947A-1/C
Sequence 1, Application US/08748947A

Patent No. 5854031

GENERAL INFORMATION:

APPLICANT: Jigami et al., Yoshifumi

TITLE OF INVENTION: MANNOSE-1-PHOSPHATE TRANSFERASE GENE

TITLE OF INVENTION: FROM YEAST, AND A PROCESS FOR PRODUCING

TITLE OF INVENTION: MANNOSE-1-PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS AND

TITLE OF INVENTION: PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS BY USE OF THE

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 601 Thirteenth Street, NW

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/748.947A

FILING DATE: 14-NOV-1996

CLASSIFICATION: A35

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 299509/1995

FILING DATE: 17-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: Ellison, Eldora L.

REGISTRATION NUMBER: 39.967

REFERENCE/DOCKET NUMBER: 08206/006001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202/783-5070

TELEFAX: 202/783-2331

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1380 base pairs

TYPE: nucleic acid

STRANDEDNESS: Single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1339

US-08-748-947A-1

Query Match 1.0%; Score 17; DB 2; Length 1380;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 560 ATTGCTAATGAATTA 576
|||||
DB 648 ATTGCTAATGAATTA 632

RESULT 21

US-09-110-959A-3/C

Sequence 3, Application US/09110959A

Patent No. 6268197

GENERAL INFORMATION:

APPLICANT: Schulein, Martin

APPLICANT: Outtrup, Helle

APPLICANT: Jorgensen, Per Lina

APPLICANT: Bjornvad, Mads Eskelund

TITLE OF INVENTION: Alkaline xyloglucanase

FILE REFERENCE: 5206.200-US

CURRENT APPLICATION NUMBER: US/09/110.959A

PRIOR FILING DATE: 1998-07-07

PRIOR APPLICATION NUMBER: 0822/97

PRIOR FILING DATE: 1997-07-07

PRIOR APPLICATION NUMBER: 1213/97

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/054,039

;; PRIOR FILING DATE: 1997-07-28
;; PRIOR APPLICATION NUMBER: 60/063,694
;; PRIOR FILING DATE: 1997-10-28
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 1614
;; TYPE: DNA
;; ORGANISM: Bacillus agaradhaerens NCIMB 40482
US-09-110-959A-3

Query Match 1.0%; Score 17; DB 4; Length 1614;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 TTTTATTAAGTCTCTTG 929
Db 144 TTTTATTAAGTCTCTTG 128

RESULT 22
US-09-270-542-101/c
; Sequence 101, Application US/09270542
; Patent No. 6322976
; GENERAL INFORMATION:
; APPLICANT: Altman, Timothy
; APPLICANT: Scott, James
; APPLICANT: Stanton, Lawrence
; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and
; TITLE OF INVENTION: Therapy
; FILE REFERENCE: 4198/78179
; CURRENT APPLICATION NUMBER: US/09/270,542
; CURRENT FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 09/221,222
; EARLIER FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 101
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-270-542-101

Query Match 1.0%; Score 17; DB 4; Length 2561;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 CCTTTTAATTTCCTT 368
Db 1444 CCTTTTAATTTCCTT 1428

RESULT 23
US-09-270-542-119/c
; Sequence 119, Application US/09270542
; Patent No. 6322976
; GENERAL INFORMATION:
; APPLICANT: Altman, Timothy
; APPLICANT: Scott, James
; APPLICANT: Stanton, Lawrence
; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and
; TITLE OF INVENTION: Therapy
; FILE REFERENCE: 4198/78179
; CURRENT APPLICATION NUMBER: US/09/270,542
; CURRENT FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 09/221,222
; EARLIER FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Rattus norvegicus

US-09-270-542-119

Query Match 1.0%; Score 17; DB 4; Length 2561;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 CCTTTTAATTTCCTT 368
Db 1444 CCTTTTAATTTCCTT 1428

RESULT 24
US-08-158-189-1
; Sequence 1, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Philip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2880 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1399..1572
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2552..2663
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 1328..1334
; FEATURE:
; NAME/KEY: CAAT_signal
; LOCATION: 1267..1271
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 1569..1576
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 2539..2549
; FEATURE:

```

; NAME/KEY: polyA_signal
; LOCATION: 2770..2775
US-08-158-189-1

```

Query Match	1.0%;	Score 17;	DB 1;	Length 2880;
Best Local Similarity	100.0%;	Pred. No. 71;		
Matches 17; Conservative	0;	Mismatches	0;	Indels

Qy	352	CCCTTTTAAATTTCTTT	368
Db	2804	CCCTTTTAAATTTCTTT	2820

RESULT 25
 US-09-360-186-2/c
 Sequence 2, Application US/09360186
 Patent No. 6262343
 GENERAL INFORMATION:
 APPLICANT: Staskawicz, et al.
 TITLE OF INVENTION: Bst Resistance Gene
 FILE REFERENCE: 50687
 CURRENT APPLICATION NUMBER: US/09/360,186
 CURRENT FILING DATE: 1999-07-23
 EARLIER APPLICATION NUMBER: 60/093,957
 EARLIER FILING DATE: 1998-07-23
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 3099
 TYPE: DNA
 ORGANISM: Capsicum annuum
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (93)..(2810)
 US-09-360-186-2

Query Match	1.0%	Score 17	DB 4	Length 3099
Best Local Similarity	100.0%	Pred. No. 71		
Matches 17; Conservative	0	Mismatches	0	Gaps 0

QY	141	TTCCCTAATGTTCTTCTG	157
Db	2813	TTCCCTAATGTTCTTCTG	2797

RESULT 26
 US-08-475-844-8
 Sequence 8, Application US/08475844
 Patent No. 5972643
 GENERAL INFORMATION:
 APPLICANT: Lobanenko, Victor V.
 APPLICANT: Nelman, Paul E.
 APPLICANT: Klenova, Elena M.
 APPLICANT: Goodwin, Graham H.
 APPLICANT: Filippova, Galina N.
 APPLICANT: Collins, Steven J.
 TITLE OF INVENTION: CTCF
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend Kourie and Crew
 STREET: One Market Place
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94105
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/475,844
 FILING DATE: 07-JUN-1995

1 CLASSIFICATION: 536
2
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: 08/261,680
5 FILING DATE: 17-JUN-1994
6
7 CLASSIFICATION: 536
8
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Parmelee, Steven W.
11 REGISTRATION NUMBER: 31,990
12
13 REFERENCE/DOCKET NUMBER: 14538A-11-1
14
15 TELECOMMUNICATION INFORMATION:

```

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

```

```

; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:

```

```

; NAME/NO: 000
; LOCATION: 292..2475
; FEATURE:

```

```

OTHER INFORMATION: /label= exon2
FEATURE:

```

```
OTHER INFORMATION: /label= exon3
FEATURE:
```

```
OTHER INFORMATION: /label= exon4
FEATURE:
```

```

OTHER INFORMATION: /label= exon5
FEATURE:
;
;
;

```

```

OTHER INFORMATION: /label= exon6
FEATURE:
;
;

```

```
OTHER INFORMATION: /label= exon7
FEATURE:
```

OTHER INFORMATION: /label= exon8
US-08-475-844-8

Query Match	1.0%	Score 17	DB 2	Length 3810
Best Local Similarity	100.0%	Pred. No. 70		
Matches 17; Conservative	0	Mismatches	0	Gaps 0

QY 360 ATTTCTTTGCCCTGTT 376
 |||||
 Db 3311 ATTTCTTTGCCCTGTT 3327

```

RESULTS
27
PCT-US95-08429-8
; Sequence 8, Application PC/TUS9508429
;
GENERAL INFORMATION:
;
APPLICANT:
;
TITLE OF INVENTION: CTCF
;
NUMBER OF SEQUENCES: 21
;
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08429
FILING DATE: 15-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/261,680
FILING DATE: 17-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-11-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: human CTCF cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 292..2475
FEATURE:
NAME/KEY: exon
LOCATION: 281..1074
OTHER INFORMATION: /label= exon2
FEATURE:
NAME/KEY: exon
LOCATION: 1075..1245
OTHER INFORMATION: /label= exon3
FEATURE:
NAME/KEY: exon
LOCATION: 1246..1379
OTHER INFORMATION: /label= exon4
FEATURE:
NAME/KEY: exon
LOCATION: 1380..1499
OTHER INFORMATION: /label= exon5
FEATURE:
NAME/KEY: exon
LOCATION: 1500..1649
OTHER INFORMATION: /label= exon6
FEATURE:
NAME/KEY: exon
LOCATION: 1650..1810
OTHER INFORMATION: /label= exon7
FEATURE:
NAME/KEY: exon
LOCATION: 1810..1992
OTHER INFORMATION: /label= exon8
PCT-US95-08429-8

Query Match 1.0%; Score 17; DB 5; Length 3810;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 360 ATTTCTTTGCCCTGTT 376
|||||
Db 3311 ATTTCTTTGCCCTGTT 3327

RESULT 28
```

```

US-09-221-235-1/c
Sequence 1, Application US/09221235
Patent No. 6043040
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,235
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4137
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: (297)..(1202)
US-09-221-235-1

Query Match 1.0%; Score 17; DB 3; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 AAAATAAAGTATTTC A 40
|||||
Db 4005 AAAATAAAGTATTTC A 3989

RESULT 29
US-09-221-928-1/c
Sequence 1, Application US/09221928
Patent No. 6121030
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,928
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4137
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: (297)..(1202)
US-09-221-928-1

Query Match 1.0%; Score 17; DB 3; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 AAAATAAAGTATTTC A 40
|||||
Db 4005 AAAATAAAGTATTTC A 3989

RESULT 30
US-09-221-527-1/c
Sequence 1, Application US/09221527
Patent No. 6146832
GENERAL INFORMATION:
APPLICANT: Acton, Susan
```



```

RESULT 32
US-09-221-416-1/c
; Sequence 1, Application US/09221416
; Patent No. 6153417
;
; GENERAL INFORMATION:
;
; APPLICANT: Acton, Susan
;
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;
; FILE REFERENCE: NMI-050
;
; CURRENT APPLICATION NUMBER: US/09/221,416
;
; CURRENT FILING DATE: 1998-12-28
;
; EARLIER APPLICATION NUMBER: 09/163,115

```

```

RESULT 34
US-09-163-115-1/C
; Sequence 1, Application US/09163115A
; Patent No. 6183962
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/163,115A
; CURRENT FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4137
; TYPE: DNA
; ORGANISM: Homo sapiens

```

FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-163-115-1

Query Match 1.0%; Score 17; DB 4; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 AAAATAAAGTATTCA 40
|||||
Db 4005 AAAATAAAGTATTCA 3989

RESULT 35
US-09-221-528-1/c
: Sequence 1, Application US/09221528
: Patent No. 6190874
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan
: TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
: FILE REFERENCE: MNI-050
: CURRENT APPLICATION NUMBER: US/09/221,528
: CURRENT FILING DATE: 1998-12-28
: EARLIER APPLICATION NUMBER: 09/163,115
: EARLIER FILING DATE: 1998-09-29
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4137
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-221-528-1

Query Match 1.0%; Score 17; DB 4; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 AAAATAAAGTATTCA 40
|||||
Db 4005 AAAATAAAGTATTCA 3989

RESULT 36
US-09-593-553-1/c
: Sequence 1, Application US/09593553
: Patent No. 6200770
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan
: TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
: FILE REFERENCE: MNI-050
: CURRENT APPLICATION NUMBER: US/09/593,553
: CURRENT FILING DATE: 2000-06-14
: PRIOR APPLICATION NUMBER: 09/163,115
: PRIOR FILING DATE: 1998-09-28
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4137
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-593-553-1

Query Match 1.0%; Score 17; DB 4; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 AAAATAAAGTATTCA 40
|||||
Db 4005 AAAATAAAGTATTCA 3989

RESULT 37
US-09-221-237-1/c
: Sequence 1, Application US/09221237
: Patent No. 6214597
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan
: TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
: FILE REFERENCE: MNI-050
: CURRENT APPLICATION NUMBER: US/09/221,237
: CURRENT FILING DATE: 1998-12-28
: EARLIER APPLICATION NUMBER: 09/163,115
: EARLIER FILING DATE: 1998-09-29
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4137
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-221-237-1

Query Match 1.0%; Score 17; DB 4; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 AAAATAAAGTATTCA 40
|||||
Db 4005 AAAATAAAGTATTCA 3989

RESULT 38
US-09-049-671-2/c
: Sequence 2, Application US/09049671
: Patent No. 5928674
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Baughn, Mariah
: TITLE OF INVENTION: NEKI-RELATED PROTEIN KINASE
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/049,671
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:

```

: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0492 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4170 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: UTRSNOR01
: CLONE: 3069734
: US-09-049-671-2.

Query Match          1.0%; Score 17; DB 2; Length 4170;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAATAAAGTATTTC 40
Db 4046 AAATAAAGTATTTC 4030

RESULT 39
US-09-295-068-2/c
: Sequence 2, Application US/09295068
: Patent No. 6030801
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Baughn, Mariah
: TITLE OF INVENTION: NEKI-RELATED PROTEIN KINASE
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESS: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/295,068
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/049,671
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0492 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4170 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: UTRSNOR01

: CLONE: 3069734
: US-09-295-068-2

Query Match          1.0%; Score 17; DB 3; Length 4170;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAATAAAGTATTTC 40
Db 4046 AAATAAAGTATTTC 4030

RESULT 40
US-08-853-310-3/c
: Sequence 3, Application US/08853310
: Patent No. 5948640
: GENERAL INFORMATION:
: APPLICANT: Randazzo, Filippo
: TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as a Tumor Suppre
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Chilton Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 94608
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/853,310
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Gutb, Joseph H.
: REGISTRATION NUMBER: 31,261
: REFERENCE/DOCKET NUMBER: 1228,003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 923-3888
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5362 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-853-310-3

Query Match          1.0%; Score 17; DB 2; Length 5362;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1466 TGCAGTATCTTCTTT 1482
Db 1207 TGCAGTATCTTCTTT 1191

Search completed: January 21, 2003, 21:46:50
Job time : 187 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 17:12:48 : Search time 2947 Seconds

(without alignments) 16166.031 Million cell updates/sec

Title: US-09-869-433-1

Sequence: 1 gaataaaactatcagaa.....gttgatccctactcgtcttt 1637

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_mu:*

19: em_mu:*

20: em_om:*

21: em_ot:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_hum:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pin:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1633.8	99.8	13389	1	AE002202	AE002202 Chlamydia
2	1632.2	99.7	11421	1	AE001619	AE001619 Chlamydia
3	1632.2	99.7	300650	1	AP002546	AP002546 Chlamydia
4	883.4	54.0	14168	1	AE001281	AE001281 Chlamydia
5	865.2	52.9	1587	1	TAJ10586	TAJ10586 Chlamydia
6	862.6	52.7	23773	1	AE002301	AE002301 Chlamydia
7	380.2	23.2	1886	6	AX098423	AX098423 Sequence
8	380.2	23.2	2300	8	STP1ATDTR	STP1ATDTR
9	374.8	22.9	2126	8	GS0251356	GS0251356
10	365	22.3	2122	8	AY098893	AY098893 Citrus hy
11	352.4	21.5	2110	8	AF428316	AF428316 Arabidops
12	352.4	21.5	2181	8	ATAMPWR	ATAMPWR
13	350.8	21.4	2151	8	AY045903	AY045903 Arabidops
14	336.6	20.6	1896	8	AY128844	AY128844 Arabidops
15	336.6	20.6	2090	8	AY081350	AY081350 Arabidops
16	336.6	20.6	2146	8	AY084374	AY084374 Arabidops
17	330	20.2	1770	6	AX098422	AX098422 Sequence
18	301.4	18.4	1823	6	AX100499	AX100499 Sequence
19	301.4	18.4	10163	1	AE001646	AE001646 Chlamydia
20	301.4	18.4	10407	1	AE002174	AE002174 Chlamydia
21	301.4	18.4	300550	1	AP002547	AP002547 Chlamydia
22	293.2	17.9	1604	1	RIRATPDTA	RIRATPDTA
23	293.2	17.9	282610	1	RPRXX01	RPRXX01
24	289.4	17.7	14117	1	AE008575	AE008575 Rickettsi
25	283.8	17.3	2139	8	ATATP2	ATATP2
26	274.2	16.8	1505	12	AF087957	AF087957 Synthetic
27	270.2	16.5	1900	1	AY120885	AY120885 Holospora
28	256.4	15.7	24297	1	AE001323	AE001323 Chlamydia
29	252.2	15.4	1623	1	TAJ10587	TAJ10587 Chlamydia
30	232.8	14.2	312430	1	RPRXX02	RPRXX02
31	227.2	13.9	12084	1	AE002346	AE002346 Chlamydia
32	217.2	13.3	10893	1	AE008630	AE008630 Rickettsi
33	215.8	13.2	10200	1	AE008626	AE008626 Rickettsi
34	214	13.1	312430	1	RPRXX02	RPRXX02
35	211.8	12.9	13627	1	AE008614	AE008614 Rickettsi
36	200.8	12.3	279110	1	RPRXX03	RPRXX03
37	199.2	12.2	237523	1	RPRXX04	RPRXX04
38	194.4	11.9	12723	1	AE008663	AE008663 Rickettsi
39	139.8	8.5	132990	8	AC018848	AC018848 Arabidops
40	138	8.4	217975	2	AC126791	AC126791 Medicago
41	136.6	8.3	174617	2	AP003977	AP003977 Oryza sat
42	134.8	8.2	156465	8	AP003234	AP003234 Oryza sat
43	134.8	8.2	172003	8	AP002910	AP002910 Oryza sat
44	132.2	8.1	91001	8	AC013453	AC013453 Arabidops
45	127.8	7.8	148018	2	AP003974	AP003974 Oryza sat

ALIGNMENTS

RESULT 1

AE002202/c

LOCUS

DEFINITION

Chlamydia pneumoniae AR39, section 33 of 94 of the complete genome.

ACCESSION

AE002202 AE002161

VERSION

AE002202.2 GI:8163425

KEYWORDS

SOURCE

ORGANISM

Chlamydia pneumoniae AR39.

REFERENCE

1 (bases 1 to 13389)

Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Unayam,L.A., Uppack,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,

Pred. No. is the number of results predicted by chance to have a

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
2 (bases 1 to 13389)
Read, T. D., Brunham, R. C., Shen, C., Gill, S. R., Heidelberg, J. F., White, O., Hickey, E. K., Peterson, J., Umayam, L. A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouli, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S. L., Eisen, J. and Fraser, C. M.
Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
20150255
10684935

TITLE
JOURNAL
COMMENT
FEATURES
Source
1. 13389
/organism="Chlamydia pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
163. 339
/gene="CP0400"
163. 339
/gene="CP0400"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38245.1"
/db_xref="gi:7189325"
/translation="MLSYMKLIPIQDSIERETDSKRKKITFTIYCSSKVLGHFFS
HLDKHNKHESIGV"
597. 1448
/gene="CP0401"
597. 1448
/gene="CP0401"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38246.1"
/db_xref="gi:7189326"
/translation="MVNIQPYRNQVNSQATQPSVQCPALSLITVSVAVALAIYA
LVCSQSLSTLGLTALVLSLIFRSAMEMLYKMKOEKELLIPKIMELIOEHPSI
VVDIIRDEVSIYEIHLISILNKTWFDKAPVLOEKLGIEKEDVHSPKLPNF
BEILLQHPLEMLRLVPMVDVTPGYGYGCGPLGLYENAPSLFERRSLLKRI
SFGFALLLEDGLKKNWSSSELVQIRQLFTRYADKEVDEALNDVDEGFDLSLHL
IFSHKLS"
1455. 1808
/gene="CP0402"
1455. 1808
/gene="CP0402"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38247.1"
/db_xref="gi:7189327"
/translation="MOLFQYMNESGMDMLCDPDSQSGFQSLRVGLHSSMALYEAK
EQFLPEVSLTMELEIMQLSKPTKHGAKDLCNVPEKHFORROTLSGLDINQRP
ENFLNPKTHLDRE"
1842. 1958
/gene="CP0403"
1842. 1958
/gene="CP0403"
/note="hypothetical protein; identified by Glimmer2; putative"

gene
CDS
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38248.1"
/db_xref="gi:7189328"
/translation="MKSLSVRSAPFGSIVKSLPNLKGKRELFEMGFITGNS"
2019. 3326
/gene="CP0404"
2019. 3326
/gene="CP0404"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38249.1"
/db_xref="gi:7189329"
/translation="MAMSKTIVPALFNKHCSTPFLSRRTVRIAIAISFCIGAL
AALGLAPVSYIVSYIAFVILSLVIALIFGEKKLPPTPIIDPRTHYIDEA
YGLSIAFVRQOVTLAEFRQFSTALCNISPEEKIKLPSELSKYSVFCISRLAGD
LEKNWPIFEEDLSQTCPLWLOKFRISAGDPOVORDLPRECYGWLGLYSTAK
ATICKETHHTLOQITKEDVILKRAKAIQEKMDDEKAIYERTITTYTAGTILKTEA
GGLKETISKELLSLHGSFSDQLITQLPROMDWALYERIVDSTAYNLCALVGA
LSQMLDESSIDEDVNLGLYIQLDKVAQAFASDEPKKELGLRLHLSVSKRL
ESVLRQGLHRIALDHGNARARYDVNFEVTGARIRKTSIFPKD"
complement(3404. 4795)
/gene="CP0405"
complement(44776. 5021)
/gene="CP0405"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38250.1"
/db_xref="gi:7189330"
/translation="MLNLRYNTQIDEPKCKMSNTSPYIQNNRSCNYFELKNTTTH
IVISAILLGALIFLCAVAPVSYILSGALLGLLALIGVILIGIKKIPMISKEQ
VEPOLVIRIRAHYPKFVDFSEKPKMLKDLISFDILNLOHSEVGSSTNYNSEL
QOKIDFEGIKARTIKNEVRTSLKLESASASRPPLSLAKILOKVPFFNIGETSIG
SKVDELHRYKIGSLBEDLSQYTRPEKPLWYLPIDOFRTNSIILNLTIVARVL
TRDVFQHLKYAALGEMVNLNSDLNTMKQDLFAFYHAAYOSKHLQSPSLQEDFYML
LICIPIKHYSKQMSLKITVPADLMENLCITLDHGRPODMEFASLIGLYDGLTH
KESEAFSLSLTLSDQFKTIRQSTNAMLNMLATHNSTFRSLPITVHPLKRSVF
SQPEDSSLLIG"
complement(44776. 5021)
/gene="CP0406"
complement(44776. 5021)
/gene="CP0406"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38251.1"
/db_xref="gi:7189331"
/translation="MRPRNIKSLIFIKRIYRSQSGEKGARPFKKSTTSVLILL
LEATFNENPSSITONNNKNNKNNISINRTFVKFTI"
complement(5081. 6361)
/gene="CP0407"
complement(5081. 6361)
/gene="CP0407"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38252.1"
/db_xref="gi:7189332"
/translation="MRHSQCMKCSPTLTPVPHFLKNDCEHSCSKIKTTARLLIG
LVIALVLSLVSFVLAAPISYATIGTALAVIILITLVALLAKSKVLPPIPELKI
ITNRYKEVFEVFKTSHLVNELKIFNCKNSGDLPPNKHAKAEFGIDILKSIDLT
LPFEFEILLQNCPLWLSHFIDKTESVAGEISGINKTKQYVGLGLPLAFHKGYTTTFH

```
SYTRPLTLTISEQYKFLYSKASKNOMDPSVKKTCEDIEIKFELPHNMLFRKDVOGISQ
FLPEFSGHITWEOAOIMOLINPDNMKLCQFDKARGCSNATGEGFINTETNEDPV
SSNLEPNTMNTMOELIKYLEKKEKESHPHSAIYOKVCVNTTHONLKRKWOVRNT
SSQWTSLSPOYAPHAQYKLEKKEISSLPIRSSL"
gene
complement(6479. .8026)
/gene="CP0408"
complement(6479. .8026)
/gene="CP0408"
GB:AJ235269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ADP, ATP carrier protein"
/protein_id="AAF38252.1"
/db_xref="GI:7189332"
/translation="MTKTEKPEFGKLSFLMHPHITHELKAVLPMLFEICITENVTL
RDYDITLIVGAPSGAEAIPIKFLWVPCAIIIMLYAKLSNLSKALDYAAGTPE
LIFPAPPTVYPLRDVLPHEFADRLQAILPGLIGLVALIRWTEFAVLAELWG
SVMLSLMWSGANEITKHEAKREVALIGANISLASGAIWASKLRAVSEGVG
PMGISLRLLMAMTIVSGVLMAKYMWIKKNVLTDPREFNPEPMOKGAKPKNMKD
SELYARSPYLLIALIVAGICINLEYWKSOKIQYNNNDYSEFNGNPSFHTG
VVSYLMLFVGNAVIRKFGWLFGALVTPVNLVIGIVFALVIFRNQASGLVAFEGTT
PLMLAVVGAIONLSKSTKYPALFDPSTEKEMAYIPLDQOKVGRAAIDVVARFGSG
GALIOGGLVYCGSIGAMTPYLAIVLFIILVIMVSATKLKMLFLAOSALKEOVAOE
DSADASS"
gene
complement(8227. .8742)
/gene="CP0409"
complement(8227. .8742)
/gene="CP0409"
/notes="conserved hypothetical protein; identified by
glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38253.1"
/db_xref="GI:7189333"
/translation="MGPMVADEOSHIKEIEIKAIKRVCKGNKENDLCRYLPSPSG
YMHFPLTKKMSAABEOLKMLKTFILSEFPRTINPDRAPRDSKKRDDINFTKTD
IERYLELAROYGDNDLLARFSPKPLTSLKRELLIRSNGLIVSELMNAVYEAAYKAVS
SPNLEVTSPFY"
gene
complement(8910. .9113)
Query Match 99.8%; Score 1633.8; DB 1; Length 13389;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 421 AATTTGCTGACCGTTTACAGGCATCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 480
|||||
DB 7656 AATTTGCTGACCGTTTACAGGCATCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 7597
QY 481 TAAGAACTGAGCATTTGCTGATTTTATGACTTGTGMACTATGAGGAGAACGCTCATGC 540
|||||
DB 7596 TAAGAACTGAGCATTTGCTGATTTTATGACTTGTGMACTATGAGGAGAACGCTCATGC 7537
QY 541 TATCTCTAATGTTCTGGGGATTGCTAATGAATTAACAAAATTCACGAGCAAAAGCGTT 600
|||||
DB 7536 TATCTCTAATGTTCTGGGGATTGCTAATGAATTAACAAAATTCACGAGCAAAAGCGTT 7477
QY 601 TCTAGCTCTTTGGGTATCGGAGCAATATTTCTTCTTCTTCTGCTGCTGCTGCAATTG 660
|||||
DB 7476 TCTAGCTCTTTGGGTATCGGAGCAATATTTCTTCTTCTTCTGCTGCTGCTGCAATTG 7417
QY 661 TTGCGGCTTCAAGTTGAGAGCTCCGTTTCTGAAGGTGTAGATCCTTGGGAAATTTCTT 720
|||||
DB 7416 TTGCGGCTTCAAGTTGAGAGCTCCGTTTCTGAAGGTGTAGATCCTTGGGAAATTTCTT 7357
QY 721 TACGCTTTTGAATGGCTATGACTATTTGATCTGACCTGTCTTATGCGCAATTACTGCT 780
|||||
DB 7356 TACGCTTTTGAATGGCTATGACTATTTGATCTGACCTGTCTTATGCGCAATTACTGCT 7297
QY 781 GGATCAATAGAAAGCATTTGAGAGCAATCCGCTGCTATATATCCAGAAAGAAAGCAAAAG 840
|||||
DB 7296 GGATCAATAGAAAGCATTTGAGAGCAATCCGCTGCTATATATCCAGAAAGAAAGCAAAAG 7237
QY 841 GGAAGAAAGGCTTAACCTAAATGAATGAAGATAGTCTCTCTCTCTTGTAGATAGAT 900
|||||
DB 7236 GGAAGAAAGGCTTAACCTAAATGAATGAAGATAGTCTCTCTCTCTTGTAGATAGAT 7177
QY 901 CTCTCTATATTTCTTTATTAATCTCTCTGTTATTTGCTATGCTATTTGCAATTAACCTAA 960
|||||
DB 7176 CTCTCTATATTTCTTTATTAATCTCTCTGTTATTTGCTATGCTATTTGCAATTAACCTAA 7117
QY 961 TCGAGCTGACTGGAAAGATCGAGTGAAGCTCAATATTCATATTCATATGATGACTATAGAG 1020
|||||
DB 7116 TCGAGCTGACTGGAAAGATCGAGTGAAGCTCAATATTCATATTCATATGATGACTATAGAG 7057
QY 1021 ACTTCATGAGGAAATCTCTCTGAGCTGGAGTATCCGTACTATCATGATTAATTTG 1080
|||||
DB 7056 ACTTCATGAGGAAATCTCTCTGAGCTGGAGTATCCGTACTATCATGATTAATTTG 6997
QY 1081 TTGTTGTTAAGCTCATTTGTAATTTGATGGTTAACTGGAGCCCTAGTACTCTGTCA 1140
|||||
DB 6996 TTGTTGTTAAGCTCATTTGTAATTTGATGGTTAACTGGAGCCCTAGTACTCTGTCA 6937
QY 1141 TGGTCTGCTAACAGTATCGTTTCTTCTGCTCTGTTATCTTTAGAAACCAAGCTTCG 1200
|||||
DB 6936 TGGTCTGCTAACAGTATCGTTTCTTCTGCTCTGTTATCTTTAGAAACCAAGCTTCG 6877
QY 1201 GGCCTGCTGCTATGTTTCGCTAGCACTCTCTATGCTAGTACTGTTGTCGAGAGTATAC 1260
|||||
DB 6876 GGCCTGCTGCTATGTTTCGCTAGCACTCTCTATGCTAGTACTGTTGTCGAGAGTATAC 6817
QY 1261 AGAATATTTCTTTCGAAATTCACAAATACGCTCTCTTGGACTCAACTAAAGAAATGGCCT 1320
|||||
DB 6816 AGAATATTTCTTTCGAAATTCACAAATACGCTCTCTTGGACTCAACTAAAGAAATGGCCT 6757
QY 1321 ATATCCCTGTCAGCAAGAGCAAAAGCAAAAGGTAAGGCTGCTATGTAAGTTGGCG 1380
|||||
DB 6756 ATATCCCTGTCAGCAAGAGCAAAAGGTAAGGCTGCTATGTAAGTTGGCG 6697
QY 1381 CCCGCTTCGAAATTCAGAGAGAGCTTATATCCAAAGAGTTGCTGCTATCTGTGNA 1440
|||||
DB 6696 CCCGCTTCGAAATTCAGAGAGAGCTTATATCCAAAGAGTTGCTGCTATCTGTGNA 6637
QY 1441 GTATTGAGCTATGACCCCTTATCTTGGAGTATCTCTTCTTTCATCATTTGCTATTTGGT 1500
|||||
DB 6636 GTATTGAGCTATGACCCCTTATCTTGGAGTATCTCTTCTTTCATCATTTGCTATTTGGT 6577
```

QY 1501 TGGTTTCGCACTAGTTAAACAACATATCTTAGCCGAGTCTGCTTTAAAGACAAG 1560
|||||
Db 6576 TGGTTTCGCACTAGTTAAACAACATATCTTAGCCGAGTCTGCTTTAAAGACAAG 6517
QY 1561 AAGTGGCAAGAAATTCAGCTCTGCTTCATAGATTGCTTACTTACTTGTT 1620
|||||
Db 6516 AAGTGGCAAGAAATTCAGCTCTGCTTCATAGATTGCTTACTTACTTGTT 6457
QY 1621 GATCCCTACCTGCTTTT 1637
|||||
Db 6456 GATCCCTACCTGCTTTT 6440
RESULT 2
AE001619 11421 bp DNA linear BCT 01-DEC-2000
LOCUS Chlamydia pneumoniae section 35 of 103 of the complete genome.
DEFINITION AE001619 AE001363
ACCESSION AE001619.1 GI:4376620
VERSION
KEYWORDS
SOURCE Chlamydia pneumoniae CML029.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 11421)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 2 (bases 1 to 11421)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
location/Qualifiers
1. 11421
/organism="Chlamydia pneumoniae CML029"
/strain="CML029"
/db_xref="taxon:115713"
complement(95..1069)
/gene="ytdg"
/note="Cpn0346"
complement(95..1069)
/gene="ytdg"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AADI8490.1"
/db_xref="GI:4376621"
/translation="MALGSPYGVSPFOFVSFRLFGSLFTGSLYTDIQLIVE
LAISGSAFGTFLVLRKMANVANASHVLPGLVYCYCLFTHOITLSLGTLLAAMA
TAMLTGLIYFIRNTEFKVSSESTALVSLSLSILYLIPMTKNAITGIELVGNAD
SLRKEDIFPTIYILANAVITTFARSLVCSFDSVFASSLGIPIRLVDTLITFQJSA
CLVGARAVGVLMALAEFLIIPSLAKVIASIRSLAMASLVISGTAFLAPASSRAIL
SAYDLGTSISGIVLPTMVIYKPIFYRGRYFSKFEKISEKSSQY"
complement(1069..2415)
/gene="ytdg"
/note="Cpn0347"
complement(1069..2415)
/gene="ytdg"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AADI8491.1"
/db_xref="GI:4376622"
/translation="MLSCVFSDFIFLSSFLAVTLICMTALMCTIILISKOPLSLSL
SHASYPGLVGLAMAOYVFSLOASIFMIVLFGCAASFGYGIIVPLGKVKLHKDSLA
CFVLVFEFALGVLAIVYKESPSPLINRIWAYLGOAATLGFLEATLAAIVPASAFA
LMMVROIYVTFDKDPAVNGCLKIVLYEALSLITSLIVSGRSVGYLISAMRYA
PSIGAROLSRISTTILLSAFGGISGALCSYISVAFTCAITIGQQAIVPTLPTGLIV
VICAGLLAGICLLFSRSGWVIRFVRKRHFSFSKDOEHLKVFHISHNLENIISVD

gene
complement(2418..3197)
/gene="ytdg_1"
/note="Cpn0348"
complement(2418..3197)
/gene="ytdg_1"
/codon_start=1
/transl_table=11
/product="rRNA methylase"
/protein_id="AADI8492.1"
/db_xref="GI:4376623"
/translation="MNVKDEFMSVHNLGVNYEHAVALYHISFSLGKSLTALGPNG
AGKSTLLKASLGLIKPSSGTVYFENQKFKVRORIAYMPQASVDMDFPMTVIDLALM
GCYSYKGMGRISDDRRFAFHILERYGLSDROKLGSLSGGQOQRAFLRALMOKA
DLYMDLFSIDMASFKTSYGVLOELRDGKKTIVYVHDLSHYRQLFDHIVLLNKLRL
ICCGPTECLNGSITFPYGCIEIIEQLTAKLSNGKQFGSC"
complement(3181..4164)
/gene="ytdg"
/note="Cpn0349"
complement(3181..4164)
/gene="ytdg"
/codon_start=1
/transl_table=11
/product="Solute Protein Binding Family"
/protein_id="AADI8493.1"
/db_xref="GI:4376624"
/translation="MDAKMGYIKVMYRIFCFVACGTTFGCTNSGPNMNSRPTIISM
NRMIHDCVEYVGNRLATAYLIKSLSDPHAYEDWKGKGDIAASAYIFCGSLBHTL
SLRHLNPNNSVRLGERLJARGAFVLEDEGQIDPHIMWDLISIMEVAIEIEVLE
KPEPWSAEFANSEIELVCMSILDSMAKQCLSTPENLRYLVSGHNAFSTYRYLAT
PEEYASGAMRSRCSIEGSLSPDAQISYRDIMAAVDYINEDVSVFPEEDTLNDALAK
IVSLKSLHLVLAQKPLYSNDVDNDFSTFKHNVCLITELGVALCEQR"
4575..5081
/gene="Cpn0350"
4575..5081
/gene="Cpn0350"
/codon_start=1
/transl_table=11
/product="C7066 hypothetical protein"
/protein_id="AADI8494.1"
/db_xref="GI:4376625"
/translation="MAVESHIKEIEIKLIGKAIKRYCGKKNEDLCRYLPFGSGYWH
HTLLKMKSAAPQLKMLKFTLLESTPTPTINPKPAPRGSKKRRDIPNFTYDIER
VLELARQVGDLDLARSPPKPLTSLKRELIRSRINGVISVELMNAVVEAVKAVSSPN
LEVTSPPV"
5282..6829
/gene="adt_1"
/note="Cpn0351"
5282..6829
/gene="adt_1"
/codon_start=1
/transl_table=11
/product="ADP/ATP Translocase"
/protein_id="AADI8495.1"
/db_xref="GI:4376626"
/translation="MTKTEEKPGKRLSPFLPIHTHELKVLPMPLMFCITENYVL
RDHKDILVAGPSGAEAIPKIFMLVLAIIIFMLIYAKLSLILSKOALFYAVYPL
LIFPALEPYIYPLRVLHPTFERADRLQALIPGLIGLIVALLNMFFAAYVLAELMG
SVLSLMFGFANEITKIHAKRPTAFGLGAGANISLASGRATVMSKLAASVSEGQ
PWGISLRLMAANTIVSGLVLMASVWINKVNLNDPREYNEEQKKGKPKPMNKD
SFLYLARSPYIILLALVIAVGICINLIEVTMKSQKLQYPMNDYSEFGNFSFMTG
VVSVLIMLFVGAVIRKFGMLTGALVTPVLLTGLVEFALVIFRNOASGLVAMFGTT
PLMLAVVGAIONITLSKSTPYALFDSSTKEMAYIPDOEOKVKKKAIDVYAARFGSG
GALIQGLVITGSGISGAMPYLLAVILLFTIATLVSAATKLKLLAQSALKEQVADQ
DSAPASS"
6968..8227
/gene="Cpn0352"
6968..8227
/gene="Cpn0352"
/codon_start=1
/transl_table=11
/product="hypothetical protein"


```
/protein_id="AAD18496.1"
/db_xref="GI:4376627"
/translation="MKCSPLTVPHIFLKNDCCHRSGLKIFRIARLIGLIVLAVS
ALSFVLAAPISVAIGSTIALAAATVILITLVALLKSKVLPIPNELOKIIVRYEK
EVEFTVTHSLIVNELKIFLIMKASGIDLPNNLKKADEAFIDILKSIDILPFEPED
ILLONCPLVWLWSHFIDKTESVAGEIKNTOKYGLGAPLAFHNGYTTIFHSYRPEL
TLISESOYKFLSKASKNOWDSVSKTCBEIEFELPHNMIFRKDVQGISQFLFELS
HGITEAOAMITOLINPDNMKMLCOFDKAGGCHSMATFGGFLNTEENMEDPVSSVEPT
VNEWTKELVLEKVESPMHPASALVOKICVNTTHONLKRKMOQVRNYSQWTS
LPOYAFHAGTYRKLKIESSLRSSL"
gene
8287..8532
/gene="Cpno353"
8287..8532
/gene="Cpno353"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18497.1"
/db_xref="GI:4376628"
/translation="MFRNKKSLIFIKIRISOSGKEOKGARGPFKKSLTSLVLL
LEAIFENFESSLIQNNFNKNKKNKINIRIFVFTI"
gene
8561..9904
/gene="Cpno354"
8561..9904
/gene="Cpno354"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18498.1"
/db_xref="GI:4376629"
/translation="MSNTSPVIONNRSNYPFLKNSTTHIIVISAILLGLALAE
CVAPSVYILSGALIGLILALIGVILIKIPIMSSKEOVPOEIVNIRAHYK
FVDFSESEAKPNLKDILSFIDILNOLISEVSNVNVSEPELOKIDFEGEIALKNE
VRKASLRBESAASSRPFLPSKILQKVEFFNGEISAGSVYELHRRKTKGSL
EEDLSDYIKPEMLPTWLTPLDFRPNSSINLHTLVARLVADVOHLKVALNGE
WNLNHDNLNMQOOLFAKHAAYOSYKRLSPQOEDEYVLLCIRKIRSNWQMSL
IKIVPADLMENTCCLTDHNGRPODMESALIGITLQGLHKESEAFVLSLTLSD
QFKTIRROSTNIAFLEMLATNHTNFSRLPRTIVPHLKRVSQPEDESSLIG"
gene
complement(9982..11283)
/gene="Cpno355"
complement(9982..11283)
/gene="Cpno355"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18499.1"
/db_xref="GI:4376630"
/translation="MSMTIVPHALFKNHCHSFPPLSSRTIVRIATLSLFCIGALAA
LGCLAPVSYVGSVLAIFAVILSLVILALIFGEKLPPIPIIDRFTVIDEAYG
LSISAFVREDOVTLAEFRQSTALCNISPEKIKOLPELSKVESGSIKSLAGDLE
KNMPIFEDLISORTCPLYWLOKPTISAGPOVCRDLGVPRECYGYWGLGISTAKAT
IFCKETHTHTIIOQLKEDYVLLKNALOEKMPDVKATVEITYTYTARQTLKNEAG
LTKETISKELLLSLHGSPDOLITOLPRDAMDLCFVNSTAVNLQDLALGALS
SONLDESSIDFVNLGLYIÖDLKEAÖVASDEPKELGKFLRLHLSVSXKRLS
VLROGLRIALEHNGNARARYDVNFVTCARITHRTSIFFKD"
BASE COUNT 3791 a 2376 c 1936 g 3318 t
ORIGIN
Query Match 99.7%; Score 1632.2; DB 1; Length 11421;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
181 TACGGATACAAAGACACTTATTGTGGAGCTCCTGTTCTGTCGACAGCAATAC 240
|||||
5412 TACGGATACAAAGACACTTATTGTGGAGCTCCTGTTCTGTCGACAGCAATAC 5471
|||||
241 CTTTCATCAAGTTTTGGCTGTGTTGCCCTGTGCTATTATCTTATTTATGCAA 300
|||||
5472 CTTTCATCAAGTTTTGGCTGTGTTGCCCTGTGCTATTATCTTATTTATGCAA 5531
|||||
301 AGCTAAGTATATTTTAAAGACAGGCGTTATTTTATGCACTGGGAACGCCCTTTAA 360
|||||
5532 AGCTAAGTATATTTTAAAGACAGGCGTTATTTTATGCACTGGGAACGCCCTTTAA 5591
|||||
361 TTTTCTTTGCCCTGTGCCGACTGTAATTTATCCGCTACGGCATTTTATCCTACAG 420
|||||
5592 TTTTCTTTGCCCTGTGCCGACTGTAATTTATCCGCTACGGCATTTTATCCTACAG 5651
|||||
421 AATTTGCTGACCGTTTACAGGCCATCTACCTCCAGATTTGCTAGACTCGTTGCCATCT 480
|||||
5652 AATTTGCTGACCGTTTACAGGCCATCTACCTCCAGATTTGCTAGACTCGTTGCCATCT 5711
|||||
481 TAAAGAACTGCACATTTGCTGATTTTATGTTACTTGCTGAACTATAGGGAACGTCATGC 540
|||||
5712 TAAAGAACTGCACATTTGCTGATTTTATGTTACTTGCTGAACTATAGGGAACGTCATGC 5771
|||||
541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATTCACAGCAAGCAAGCGTT 600
|||||
5772 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATTCACAGCAAGCAAGCGTT 5831
|||||
601 TCTACGCTCTCTTTCGATACGAGCTAATATTTCTTACTAGCTCTGTCGTGCAATTG 660
|||||
5832 TCTACGCTCTCTTTCGATACGAGCTAATATTTCTTACTAGCTCTGTCGTGCAATTG 5891
|||||
661 TTTGGGCTTCAAGATTTGAGAGCTTCGCTTCTGAAAGGTGAGATCCTTGGGAATTTCTT 720
|||||
5892 TTTGGGCTTCAAGATTTGAGAGCTTCGCTTCTGAAAGGTGAGATCCTTGGGAATTTCTT 5951
|||||
721 TACGCTCTTTCGATACGATATTGATTCGAGCTTCTTTCGAGCACTTACTGCT 780
|||||
5952 TACGCTCTTTCGATACGATATTGATTCGAGCTTCTTTCGAGCACTTACTGCT 6011
|||||
781 GGATCAATTAAGAACGTAATTAAGCAGATCTCGCTTCTATTAATCAGAAAGCAAAAGC 840
|||||
6012 GGATCAATTAAGAACGTAATTAAGCAGATCTCGCTTCTATTAATCAGAAAGCAAAAGC 6071
|||||
841 GGAAGAAAGGCTCTAAGCTTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 900
|||||
6072 GGAAGAAAGGCTCTAAGCTTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 6131
|||||
901 CTCCTTAATCTCTTATTAATCACTCTCTGATTAAGCCTATGCTATGCTATTAATCACTTA 960
|||||
6132 CTCCTTAATCTCTTATTAATCACTCTCTGATTAAGCCTATGCTATGCTATTAATCACTTA 6191
|||||
961 TCGAAGTGACTTGAAGAACTGCAAGCTCAATATCTCAATATCTCAATATCTCAATATCTAGTG 1020
|||||
6192 TCGAAGTGACTTGAAGAACTGCAAGCTCAATATCTCAATATCTCAATATCTCAATATCTAGTG 6251
|||||
1021 AGTTATGAGGAACTTCTCTCTGAGCTGGGCTAGTATCGCTACTTATCATGCTATTTTG 1080
|||||
6252 AGTTATGAGGAACTTCTCTCTGAGCTGGGCTAGTATCGCTACTTATCATGCTATTTTG 6311
|||||
1081 TTGGTGCTAGACGTCATGCTCAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1140
|||||
6312 TTGGTGCTAGACGTCATGCTCAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 6371
|||||
1141 TGGTCTCTCTAAGAGTATGCTTTCTGCGCTTGTATCTTATGAAACCAAGCTTCTG 1200
|||||
6372 TGGTCTCTCTAAGAGTATGCTTTCTGCGCTTGTATCTTATGAAACCAAGCTTCTG 6431
|||||
1201 GGGTGGTGCTATGTTGGTACAACTCTCTCATGCTAGCTGTGTTGGAGGCTATAC 1260
|||||
6432 GGGTGGTGCTATGTTGGTACAACTCTCTCATGCTAGCTGTGTTGGAGGCTATAC 6491
|||||
```

QY 1261 AGAATATTTCTTTCGAATTCACAAATAATACGCTCTCTTGTACTCAACTAAGAATGGCCT 1320
 |||||||
 Db 6492 AGAATATTTCTTTCGAATTCACAAATAATACGCTCTCTTGTACTCAACTAAGAATGGCCT 6551
 |||||||
 QY 1321 ATATCCCTCTTGACCAAGCAAGAAAGTCAAGGTAAAGGTGCTATTATGATGTTGCCG 1380
 |||||||
 Db 6552 ATATCCCTCTTGACCAAGCAAGAAAGTCAAGGTAAAGGTGCTATTATGATGTTGCCG 6611
 |||||||
 QY 1381 CCCGCTTGGAATAATCAGAGAGAGCTTTAATTCACAAGAGTTTGGTCGTTATCTGTGAA 1440
 |||||||
 Db 6612 CCCGCTTGGAATAATCAGAGAGAGCTTTAATTCACAAGAGTTTGGTCGTTATCTGTGAA 6671
 |||||||
 QY 1441 GTATTGAGCTATGACCCCTTATCTTGAGATGATTTCTTTCATCATTCGTAATTTGGT 1500
 |||||||
 Db 6672 GTATTGAGCTATGACCCCTTATCTTGAGATGATTTCTTTCATCATTCGTAATTTGGT 6731
 |||||||
 QY 1501 TGGTTTTCGCACTAAGTTAAACAACACTATTCTTAGCGCAGTCTGCTTTAAGACAAG 1560
 |||||||
 Db 6732 TGGTTTTCGCACTAAGTTAAACAACACTATTCTTAGCGCAGTCTGCTTTAAGACAAG 6791
 |||||||
 QY 1561 AAGTCGCTCAAGAAGATTACAGCTCTGCTCTTCATAGAGTTGCTTCTTACTCTTGT 1620
 |||||||
 Db 6792 AAGTCGCTCAAGAAGATTACAGCTCTGCTCTTCATAGAGTTGCTTCTTACTCTTGT 6851
 |||||||
 QY 1621 GATCCCTACCTGCTTTT 1637
 |||||||
 Db 6852 GATCCCTACCTGCTTTT 6868
 |||||||

RESULT 3
 AP002546
 LOCUS
 DEFINITION
 Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,
 section 2/4.
 ACCESSION
 AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801
 AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808
 AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815
 AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA000008
 AP002546.2 GI:10176692

VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Chlamydomophila pneumoniae J138 (strain:J138) DNA.
 Chlamydomophila pneumoniae J138
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.

REFERENCE
 AUTHORS
 1
 Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,
 Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,
 Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,
 Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.
 Comparison of outer membrane protein genes omp and pmp in the whole
 genome sequences of Chlamydia pneumoniae isolates from Japan and
 the United States
 J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

2
 Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
 Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
 Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CML029 from USA
 Nucleic Acids Res. 28 (12), 2311-2314 (2000)

3
 (bases 1 to 300650)
 Shirai,M.
 Direct Submission
 Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
 School of Medicine, Department of Microbiology; 1-1-1
 Minamikogushi, Ube, Yamaguchi 755-8505, Japan
 (E-mail:shirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
 Fax:81-836-22-2415)

On or before Sep 15, 2000 this sequence version replaced
 gi:6172290, gi:6172292, gi:6172294, gi:6172296, gi:6172326,
 gi:6172338, gi:6172330, gi:6172332, gi:6172334, gi:6172336,
 gi:6172338, gi:6172382, gi:6172384, gi:6172386, gi:6172388,
 gi:6172390, gi:6172392, gi:6172394, gi:6174666, gi:6174668,

FEATURES
 source
 gi:6635174, gi:6635176, gi:6635178, gi:6635180, gi:8547433,
 gi:8547438, gi:8978640,
 Location/Qualifiers
 1..300650
 /organism="Chlamydomophila pneumoniae J138"
 /strain="J138"
 /db_xref="taxon:138677"
 /note="synonym:Chlamydia pneumoniae (strain J138)"
 144..935
 /gene="CPj0267"
 144..935
 /gene="CPj0267"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="BAA8477.1"
 /db_xref="GI:8978641"
 /translation="MSIMSLSKNTNALLNDEPAVCLNAMPDKYIINDRKTFACVTLL
 VIATMLITGVIVLLAMGSPGLSVLTIGTSVTLGTFILGIVLKIKSLAMI
 OYKRYFOEVKOKYKPEPSIPKNDVNHKLTGSLPDLIESPSPAPVAKRIACSG
 VAIIVGVTLLIGAVSVFPCGYGLALCVGFACLTGALFVGLAGLRTHSLAAGIM
 KLYITLYLSSALBERNETVADQREINITYLRECRQKRKALLE"
 937..1343
 /gene="CPj0268"
 937..1343
 /gene="CPj0268"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="BAA8478.1"
 /db_xref="GI:8978642"
 /translation="MSQCSSTSTWEMMKSFVPMKNPPTPLSPISEDEFILAEV
 FVLKPTDENAOANPCTSTPNVNGIDDLNPLLGQNEQNNANNPCTSGSNPTSLPA
 PERLPTEENSOEEPOCSNNEDIG"
 complement(1495..2472)
 /gene="CPj0269"
 complement(1495..2472)
 /gene="CPj0269"
 /codon_start=1
 /transl_table=1
 /product="dipeptidase"
 /protein_id="BAA8479.1"
 /db_xref="GI:8978643"
 /translation="MTIDMHCDLISHPHCRKDPVRCSPEDLISGVROQCAIFVP
 HSRGPNCDKONSLEFSLPNQYDGLIGLSEENSSSQKSLISRTSEMSALMD
 DTPAPICGLIKLHITKQGLAYIGTSGKDNRRGCTEAPKLSNKGVLVIMDEL
 GVPDLSHCSDKLAEDLDYTDAPKPLMAYIASISNRSRLDHRNIVDHAKEIYR
 KGIYGLNLVRSYVGDSDGLEKHLHAEENGLISSIVLGSDFYANEDENFEFNECSS
 AEAHPIVNLQIHRIFFSGKAESILSRAEKFLLQVIVEQVNPKITDVKL"
 complement(2487..3347)
 /gene="ywlC"
 complement(2487..3347)
 /gene="ywlC"
 /codon_start=1
 /transl_table=1
 /product="SuA5 superfamily-related protein"
 /protein_id="BAA8480.1"
 /db_xref="GI:8978644"
 /translation="MPDKKAQIFSLPEVMSAIHQCKIVALPTDVTYGVFLSLXASA
 EERIVALKRPSKAFALVNSIEDINISGYPLSPAKKLAQDFGATLVVKNRP
 RPKETLAFRIVDHSVREIVDHCGLTIGTSANLSEPSALTAQEIFADPADDLCTE
 DCPCHGLESTVAVSDPLIYRGLISRSVIENIAGTEAKIFRTHSAFSAFKHKIYIV
 KNOEOLVSLFSGSLDFKGVVCEHPKPKNFYTRLEALKKTKPSIVFIYDINTSDYDEL
 FPLSPYYIE"
 3556..4387
 /gene="CPj0271"
 3556..4387
 /gene="CPj0271"
 /codon_start=1
 /transl_table=1
 /product="lysophospholipase esterase"
 /protein_id="BAA8481.1"
 /db_xref="GI:8978645"

```

gene
CDS
    /translation="MTDYSFERRKIGNIEAIECPNPDPPIIILCHGYGSLADNLTF
    PSICSFSLRPTWIFPNGLIPLLENDPFSRACPLNVLLOELSLVLAQSNLOEKY
    DELFDVLETKALEELLINLNRPYNEIIIGGFSQGLATLHVLVTSONYAGALLF
    AGARLNGNGEGLKCOQVFPLOSHGDEDELLPYHGAHLNDLLLTILNQPSFSG
    GHEIPSVFQKMQVTPNWDIPARC"
    complement(4359..5255)
    /gene="dnaX_2"
    complement(4359..5255)
    /gene="dnaX_2"
    /codon_start=1
    /transl_table=11
    /product="DNA polymerase III gamma and tau"
    /protein_id="BA98482.1"
    /db_xref="GI:8978646"
    /translation="MHLSEENOGWEALLRKVYHDEVPAILLHGFTLPVLDKKAOLA
    SEILSSSPGSEHKYSOKIHPDIYOFPEPGRGRLSIDLPGRICKQIYISPEANYKI
    YIIEHADMTLAAISAFLEKEEPKHAVALITLTKAVORLRTIILSRSLSTFIEGK
    ILCSEFTSYLFERYAOCEIPLYEVSQIIKESSEDDKQVLRDQVRFMEVLLLEADRY
    TLNLGLKASALNPEHVEKEIILQLPLLPDKVLLIVESACRSILNNSAASVLEWVAIO
    LVSLOYKKEELVSPGQDLN"
    complement(5246..5866)
    /gene="tdk"
    complement(5246..5866)
    /codon_start=1
    /transl_table=11
    /product="chymidylate kinase"
    /protein_id="BA98483.1"
    /db_xref="GI:8978647"
    /translation="MFIVIEGEGSGSKSLAKALGDQLVADRKVLLTREFGCLLIGE
    RLRLDLEPHELESCCELEFLGSRQHOIEVLIIPALRGYIYICRPHDSTIVYG
    GIAGLGADVPADLCSKYVGPPLPLPNVLDDIPADIGLGRKHROKFDKFEKKPLS
    YHNRIRGEFLSLASDPSRYLVLDARESLSLIDKVMHLTQGLCT"
    complement(5871..8375)
    /gene="gyrA_1"
    complement(5871..8375)
    /gene="gyrA_1"
    /codon_start=1
    /transl_table=11
    /product="DNA gyrase subunit A"
    /protein_id="BA98484.1"
    /db_xref="GI:8978648"
    /translation="MFNKDEIIVPKLNEEMKESYLRSMSVYLSRALPDIRDGLKPS
    ORVLYAMKOLSLSPGAKHRCOKICGTSQSDYHSGVSPYTLVRAQWMMAMKYP
    VDGNGESIDGDEPAARYTEARLTSHAMLMEDLDQTDIVNPEYETHEPVPVP
    SKPMLICNSSSGTAAVGMATNIPRNHLELLIALLANQASDELQVMPGDFP
    TGGIICGSEGRSTTTTGKIKIVARLHVEDNEKHESIITTEPYNVKSRIEIO
    IANLVNKTLAGISDVROSDKDGIRVLEIKGSESIILNRLKYTDVQVTEGAM
    LADKNLPTMSIHMIISAWIRHREKVIIRRTRELNKAETRAHYLEGYLALSCDA
    LAVKIRSEGNKEHAKERIIESGFTEPOLAILLELTYQLGLEAEKIOKEYEELNK
    IAYKQVYSDGLVKDIIIRNELODLKHKVARRTTIEFDADDIRDIEDITINESVLI
    TISGDYVKKRMPVYKFEQRRGGHGVTFDKKGAFLKAYSAFTKYYLILFTNEGQ
    CYLKWOLPGEERRAKGPIINLLEGIIPGEBELAILININFDAGLFLATKRGV
    KRVSLDAFNSPRKGIKRALIEDEGDLIAACHIVSDEKVMLEFHLGNAVERPHEKRV
    PMGRTARGVNSLKNEDKVVSCQIVTENSQVLCVDCGFGKSLVDPEFTNNGYV
    GVRSLIINERNGVNLGAIPVTDHDSIILMSOGQAIRINMDVRYMGRSSTOGVRLVHL
    KEGDAIVMEKISNENDEVLVSGSEECSDPLSLR"
    complement(8391..10808)
    /gene="gyrB_1"
    complement(8391..10808)
    /gene="gyrB_1"
    /codon_start=1
    /transl_table=11
    /product="DNA gyrase subunit B"
    /protein_id="BA98485.1"
    /db_xref="GI:8978649"
    /translation="MDPKENYNDASATVLEGLQAVRERPGMTIGDITGLHLHLYE
    VVNSIDAMAGYCSRIDVRILEDGIVIVNGRGIPIEVEHERESAKRGREYSALEVV
    LTVLAGKPKDYSKVGSGLHGVCVNLSEKLVATVFKDKCKQYMEFSRGIPVPT
    PLOYVSVDRGGTETVYVDPDKIPSTCTGDESIIMKLRLBELAFLNIGRTIYFEDDRD
    SPDKYTFYEGCIOSFVSYLWONKESLSELYICGIRVGDGELTFEPALWQNSGIS
    ELVSYANNIPTRGGTHLTGFTALTVIVINTYIKAHNLANNKALIGEDIELELTA
    VISVVPNDPDEGQTKOKIGNSDVSSVAQOVVVEGALITFPEENPDIAMRIVDKFVAA

```

```

OAREAKARELTLRKSLDSARLPGKLIIDLCEKPEKCEMYIEVGSAGSAAKQGRD
RFOAILIRPTLILVEKARLOKIPONOEIGITIALCGGAGDNPISKLRYRILII
MWDADVDSHRIITLLEFFRYHMTLIEBVCYVLAPOPLVYAGKDFRRLTISEKEMD
SYLLMLGTNESIILFKSTERELRGALSFIVNIIIDVESFINTLEKKAIPSEFLEM
KEGIGYPLYLATPATGMQGRYLYSDSEKALAEETHKRIIELYVAFAVDLQNO
LKEVGLDISYLIPOKNEIVIGNEDESPSCNVCYCLEVIVNKLKMGKGEIQRKYG
LGENADQLMDTMMPEQRTLIHVSILKDAVEADHIFPLMGEVEVPPREFIESHALSI
RIINNDI"
    complement(10812..11159)
    /gene="CpJ0276"
    complement(10812..11159)
    /gene="CpJ0276"
    CDS
    Query Match 99.7%; Score 1632.2; DB 1; Length 300650;
    Best Local Similarity 99.8%; Pred. No. 0;
    Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
    1 GAAATAAAAAACATCAGATAGAAAATAAAAGTATTCAGAGGTAATATGACAAAAA 60
    93946 GAAATAAAAAACATCAGATAGAAAATAAAAGTATTCAGAGGTAATATGACAAAAA 94005
    61 CCGAAGAAAACCTTTTGGAAAAATTCGCTCTTTCTTGCGCGATACATCTACGAGC 120
    94006 CCGAAGAAAACCTTTTGGAAAAATTCGCTCTTTCTTGCGCGATACATCTACGAGC 94065
    121 TAAAGAAAGTTCGCGCAATGTCCTAATGTCCTGCTGCTATTCATTAATACGCTGT 180
    94066 TAAAGAAAGTTCGCGCAATGTCCTAATGTCCTGCTGCTATTCATTAATACGCTGT 94125
    181 TACGCGATCAAAAGACACTCTTATGTGGAGACTCTGTTCTGTGCGAGAGCAATAC 240
    94126 TACGCGATCAAAAGACACTCTTATGTGGAGACTCTGTTCTGTGCGAGAGCAATAC 94185
    241 CTTTCATCAAGTTTGGCTTTGTTGCCCTGTCGCTATATATCTTATATATGCAA 300
    94186 CTTTCATCAAGTTTGGCTTTGTTGCCCTGTCGCTATATATCTTATATATGCAA 94245
    301 ACCTAAGTAAATATTTTAACTAAGCAGCGCTTATTTATGACGAGGGAAGCCCTTTTAA 360
    94246 ACCTAAGTAAATATTTTAACTAAGCAGCGCTTATTTATGACGAGGGAAGCCCTTTTAA 94305
    361 TTTTCTTTCGCCCTGTTCCGACGTAAATTTATCCGCTACGCGATGTTTACATCTACAG 420
    94306 TTTTCTTTCGCCCTGTTCCGACGTAAATTTATCCGCTACGCGATGTTTACATCTACAG 94365
    421 AATTGCTACGCTTTACAGCCCATCTACCTCAGAGATTGCTAGAGACTGTTGCCATCT 480
    94366 AATTGCTACGCTTTACAGCCCATCTACCTCAGAGATTGCTAGAGACTGTTGCCATCT 94425
    481 TAAGAAACGAGCATTTGTCGATTTATGTCCTGCAAGCATGAGGGAAGGTCATGC 540
    94426 TAAGAAACGAGCATTTGTCGATTTATGTCCTGCAAGCATGAGGGAAGGTCATGC 94485
    541 TATCTCTAATGTTCTGGGATTTGCTAATGAATATCAAAAAATCAAGAGCAAAAGCGTT 600
    94486 TATCTCTAATGTTCTGGGATTTGCTAATGAATATCAAAAAATCAAGAGCAAAAGCGTT 94545
    601 TCTACGCTTTTTCGCTATCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTG 660
    94546 TCTACGCTTTTTCGCTATCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTG 94605
    661 TTTGGGCTCAAAAGTGAAGGCTCGCTTTCGAAAGGTGTAACATCTTGGGGAATTTCTT 720
    94606 TTTGGGCTCAAAAGTGAAGGCTCGCTTTCGAAAGGTGTAACATCTTGGGGAATTTCTT 94665
    721 TACGCTTTTTCGCTATGCTATGATTCGAGCTTCTTATGAGCACTTACTAGCT 780
    94666 TACGCTTTTTCGCTATGCTATGATTCGAGCTTCTTATGAGCACTTACTAGCT 94725
    781 GGATCAATAGAAACGATTTGACCGATCTCGCTTCTATATCCAGAAAGAAATGCAAAAGC 840
    94726 GGATCAATAGAAACGATTTGACCGATCTCGCTTCTATATCCAGAAAGAAATGCAAAAGC 94785

```

```

Qy 841 GGAAGAAAGGCTCTAAACCTAAATGATGAAAGATAGCTTCTCTATCTTGTATAGAT 900
    |||||||
Db 94786 GGAAGAAAGGCTCTAAACCTAAATGATGAAAGATAGCTTCTCTATCTTGTATAGAT 94845
Qy 901 CTCCTATATCTCTTATTAATCTCTCTGTTATGCGCATGGTATTTGCAATTAACCTAA 960
    |||||||
Db 94846 CTCCTATATCTCTTATTAATCTCTCTGTTATGCGCATGGTATTTGCAATTAACCTAA 94905
Qy 961 TCGAAGTACTTGGAAAGTACAGTGAAGTCAATATCCAAATATGATGATGACTATAGT 1020
    |||||||
Db 94906 TCGAAGTACTTGGAAAGTACAGTGAAGTCAATATCCAAATATGATGATGACTATAGT 94965
Qy 1021 AGTTATGATGGGAACCTCTCTCTGTTATGCGCATGGTATTTGCAATTAACCTAA 1080
    |||||||
Db 94966 AGTTATGATGGGAACCTCTCTCTGTTATGCGCATGGTATTTGCAATTAACCTAA 95025
Qy 1081 TTGGTGGTAAAGTCAATGCTAAATTTGATGGTAACTGGAGCCCTACTGCTCTGCA 1140
    |||||||
Db 95026 TTGGTGGTAAAGTCAATGCTAAATTTGATGGTAACTGGAGCCCTACTGCTCTGCA 95085
Qy 1141 TGGTTCTCTCAACAGGTATCGTTTTCTTCGCTCTTGTATCTTGAAGAACAGCTCTG 1200
    |||||||
Db 95086 TGGTTCTCTCAACAGGTATCGTTTTCTTCGCTCTTGTATCTTGAAGAACAGCTCTG 95145
Qy 1201 GCGTGGTGGTATGTTGGGTCACACTCTCTCATGCTGAGTGGTGTGCGAGCTATAC 1260
    |||||||
Db 95146 GCGTGGTGGTATGTTGGGTCACACTCTCTCATGCTGAGTGGTGTGCGAGCTATAC 95205
Qy 1261 AGAATATCTCTTGCAGAAATCCAGAAATATGCGCTCTTGCATCAACTAAAGAAATGGCCT 1320
    |||||||
Db 95206 AGAATATCTCTTGCAGAAATCCAGAAATATGCGCTCTTGCATCAACTAAAGAAATGGCCT 95265
Qy 1321 ATATCCCTCTTGACCAAGAGCAAGAAAGTCAAGAGTAAAGCTATGATGATTTGCCG 1380
    |||||||
Db 95266 ATATCCCTCTTGACCAAGAGCAAGAAAGTCAAGAGTAAAGCTATGATGATTTGCCG 95325
Qy 1381 CCCCCTTGGAAATTCAGAGAGAGCTTTAATCCAAAGAGTTTGGTCTGTTATCTGTGAA 1440
    |||||||
Db 95326 CCCCCTTGGAAATTCAGAGAGAGCTTTAATCCAAAGAGTTTGGTCTGTTATCTGTGAA 95385
Qy 1441 GTATTGAGATGATGACCCCTTATCTTGCAGTATCTTCTTATCATCTGCTATTTGGT 1500
    |||||||
Db 95386 GTATTGAGATGATGACCCCTTATCTTGCAGTATCTTCTTATCATCTGCTATTTGGT 95445
Qy 1501 TGGTTTTCGCACTAAGTTAAACAACAATCTTAAAGCAGCTGCTCTTAAGAACAAG 1560
    |||||||
Db 95446 TGGTTTTCGCACTAAGTTAAACAACAATCTTAAAGCAGCTGCTCTTAAGAACAAG 95505
Qy 1561 AAGTGGCTCAAGAGATTCAGCTGCTGCTTCTCATAGAGTTGCTTCTTACTTGT 1620
    |||||||
Db 95506 AAGTGGCTCAAGAGATTCAGCTGCTGCTTCTCATAGAGTTGCTTCTTACTTGT 95565
Qy 1621 GATCCCTACCTGCTTTT 1637
    |||||||
Db 95566 GATCCCTACCTGCTTTT 95582

```

```

RESULT 4
AE001281/c 14168 bp DNA linear BCT 30-OCT-2000
LOCUS Chlamydia trachomatis section 8 of 87 of the complete genome.
DEFINITION Chlamydia trachomatis section 8 of 87 of the complete genome.
ACCESSION AE001281 AE001273
VERSION AE001281.1 GI:3328454
KEYWORDS
SOURCE
ORGANISM Chlamydia trachomatis.
REFERENCE 1 (bases 1 to 14168)
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aaravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
TITLE Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis

```

```

JOURNAL Science 282 (5389), 754-759 (1998)
MEDLINE 9900809
PUBMED 9784136
REFERENCE 2 (bases 1 to 14168)
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 9920606
PUBMED 10192388
REFERENCE 3 (bases 1 to 14168)
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aaravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
TITLE Direct Submision
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
source
1. 14168
/organism="Chlamydia trachomatis"
/strain="D/DW-3/CX"
/db_xref="taxon:813"
86..847
/gene="flaA"
/note="CT061"
86..847
/gene="flaA"
/codon_start=1
/transl_table=1
/product="Sigma-28/Whig Family"
/protein_id="AAC67652.1"
/db_xref="GI:3328455"
/transl_table="MKHDLADLMQLYWSTKEIHHRDVLESYLPVKVNAHRLASGM
PSHKMEDLALGVELIRAVERPDPKSRPESYAFIKAIITGRLKRDWPRSV
YQANRLADAMDLSRLQTLGKPEPDLKCAINISQDLSPFSSRPALSLSDDS
CODDDESLAEERLADERETGVDYIKKKAISITLLELALDEKERQVNAIYYDDL
VKEIGKILGVESRSVSHSKALKLRGLSL"
1006..2244
/gene="lysS"
/note="CT062"
1006..2244
/gene="lysS"
/codon_start=1
/transl_table=1
/product="tyrosyl tRNA synthetase"
/protein_id="AAC67653.1"
/db_xref="GI:3328456"
/transl_table="MOGLDNLKRGITLNDSSAGLESITYPVSAIYGFDPPLSHIG
HWIGICFLRLAAVGIPTVALVGATGEMIDPSGKSYERSLDDQAVLDNSKRTAAL
ASYLPGRIVNADMDLGLSMVDPLRDVGNHFRIGSLMADVVKORYSDEGISTYEF
SYLLGSDYDFAHLFEKHNVLVOCGSDQMCNITSGIDYIRRRIGQAYGLYPLTDS
KGRKIGTESGTITMDPLTPPELPPYELPPYFLRDPDQSKWRTLLTLDNEITFALDR
LTSDDPAVKKYVIAEVIVKDVHSGEGLAQAATAESPASKSKSITTEALVAVESCG
VKVARADLIGKRMIDIVELGFCSSRGQARRLLIQKRLTYNDEPLADEOSILDGTOLC
FDRYVLSQGRKKQYIDIN"
2254..3696
/gene="gnd"
/note="CT063"
2254..3696
/gene="gnd"
/codon_start=1
/transl_table=1
/product="6-Phosphogluconate Dehydrogenase"
/protein_id="AAC67654.1"
/db_xref="GI:3328457"
/transl_table="MAPNTDIGLIGLAWGKNLVLNMWDHGSVSYVNSPAKTEEPL
KDHESGALOGFTTQIEVOSLKRPRKIMIMIRAGAPVEMIASLPLFEEGLILDG
GNSYVLDSEQRVYDLKKEGILFVGMVSGGSEBAGRPSIMPGNIDAMPALPIDOS
IAOVQGRPCCSWIGTGAGHFKAVANGHTEYDIDICETVETILRLNLSLEOIGN
IFPEMNQDINSYILGAGAAVLTAKDNGNAISTIDVAGOKGCRWVAEDAKVACG
PMSLIIESVARTLSTWKEVRTAADEPFIPLCCPPQDASAFIEDVREBALYAAKI
STVAGFMRLKQVSQDGMWDNLGELALIMWGCTIQSAFLDKTHQGFENSPHNSLIL
QDYFKRVLPDSETFRRRAVYLHAIGSVAIPCLSSALSIFYDGYRTVDSLSLVLQGLRDY

```

```

gene      FGANGERRDCRGEGYHTDMLTKKTRRV
           complement(3757. .5565)
           /gene="lepa"
           /note="CT064"
           complement(3757. .5565)
CDS       /gene="lepa"
           /transl_table=1
           /codon_start=1
           /product="GTPase"
           /protein_id="AAC67655.1"
           /db_xref="GI:3328458"
           /translacion="MKPKIENIRNFSIIAHIDHCKSTIADRLLESTSTIEQRENEQ
           LLDSDMLEREGITTKAHVPMTYEYEGEYELNIIDPGHDFSEYRSIIACEGA
           LLIYDAAGVOAOSLANVYLAERDEIIPVNLIDLPAAOPEAKKOIEEIGDTS
           NTIACSAKTGGIPELIIISIIIRLVPPPKPOPELTKALIPDSHPDYGVYIVVIS
           GEIKKDRITRMATKSSFEVVGIGAFIPEATLMGSLRACGYGFIANKKVKYKI
           GDTVTYVHPAKPELEGRKEIKPVFACIYPIDSSDFDTLKDALGRLOJNDALTEQ
           ENSHSLGFGFRGFLGLHLEIIFERISREDLDIATAPSVIYKVLKNGKTFIDN
           PTAYDPDALIEHEEPWVHNIIPQEYLSNMSLCMDKRGICLKTMDLDRILVLY
           ELPLNEIYSDNDKLSVTKGSGFDYRGOKKCAIKLEILINDEAVDAFSCLVHR
           DKAESKRSICEKIVDVIIPOLFKIPIQAIINKKIARETIRALAKNVYACYGDDIT
           RKRKIMDKKKKKMKKEFKYSIPNTAFEVLKME"
           complement(5751. .7337)
gene      complement(5751. .7337)
           /gene="CT065"
           complement(5751. .7337)
CDS       /gene="CT065"
           /codon_start=1
           /transl_table=1
           /product="ADP/ATP Translocase"
           /protein_id="AAC67656.1"
           /db_xref="GI:3328459"
           /translacion="MTQTAKEPFKGLRSEFLPIIHHELKVLPMLFECTSPNTIL
           RDTKDTLIVTAPGSAEAIPIKIMLVPSAVPEMLIYAKISNLINKALPRAVISP
           VVPRFLPVVYIPCRHIIHPFAFDLTOSILPSGMEFITAMLRNMTAFVYVLSLWG
           SVMSLMEWFRANETIKISEAKREFALGVANVALLISGPAILWSSKLRSLGVD
           PMGVSLYFLMAFLCSCAIIACQYMMNRVYLDPRFNPALAKSKPKMSGESF
           SYLLRSPYMLLALILYCGICINLVEYTKSOLMKOPPNPDYSAFGNFSFMTGV
           SVFVLFIGNVYIRREFGWLGTALVPIWLVGTAFVIFRDAHTGALVAGTITPL
           MVAVVGAIQNLISKSTKALFDATKEMAXIPLDOECKVKGAADIVAAAPRGKGS
           LIQGLIYVCGSIGMTPLFVALVAIITMWLTSLTKLKLFLAASAKEDBLAEEAA
           AEKESSAAKSAAPALEGS"
           complement(7688. .8164)
gene      complement(7688. .8164)
           /gene="CT066"
           complement(7688. .8164)
CDS       /gene="CT066"
           /codon_start=1
           /transl_table=1
           /product="hypothetical protein"
           /protein_id="AAC67657.1"
           /db_xref="GI:3328460"
           /translacion="MATAQITIOEITIOELITKAIKSGSKENDICRYLRPRTGYMH
           HPTLKMKNSAPOLKMLKTFVIDSTIPRSIPPRAPRGSKRRDRVNFXTIER
           VLELROVGDDLDARFSPKRPPLSLKRELIIRSNITVSTELMAVVEALSQPE"
           8822. .9802
gene      /gene="ytga"
           /note="CT067"
           8822. .9802
CDS       /gene="ytga"
           /codon_start=1
           /transl_table=1
           /product="Solute Protein Binding Family"
           /protein_id="AAC67658.1"
           /db_xref="GI:3328461"
           /translacion="WSFHTKRYKILRLGLCLACGLMNSCSSRGNOADESTIYL
           SMNRMICDSVSRITGDRVKNIVLIDGAIIDPHSYEWDGDEDRMAASQILFNGCLIEH
           SASLRKLEHNGKVDYDGLRLINKNCPDLSEEGPPDHITMDMVAQAAVYEMAAAL
           IQOFQYEDERKNADQIILESEELDRMAASLSITIPKRNVLVYHNAFVFTRRYL
           SSDAERVSQKIRAKSPLYSDNVGNVFFSTPQNVNTITTEELGTYLE"
           9795. .10574
gene      /gene="ytga"
           /note="CT068"
           9795. .10574
CDS

```

```

/gene="ytga.1"
/codon_start=1
/transl_table=1
/product="rRNA methylase"
/protein_id="AAC67659.1"
/db_xref="GI:3328462"
/translacion="MNRDNLAMSVEDLCVNVHSDVLCIHTFSLPAGAMAIIGPNC
AGSKSLKASGLIRASSGSLFSGDRFSKYHNRJAAVYPOASVMDPMTVLDLYLM
GCYKGTGNRSTDDROFARKITERVGLBAPANKOIKLSGGQOQRAFARSLMOKA
DLYLMDLEFSADMASYOMVVDVLOELSEGTIVYIHHDSLNSWRKLPDHYILNKLH
VCSGSECELTKEALFOAYGCELDLDYLTKLRSKYGSC"
10575. .11930
gene      /gene="ytgc"
           /note="CT069"
           10575. .11930
CDS       /gene="ytgc"
           /codon_start=1
           /transl_table=1
           /product="Integral Membrane Protein"
           /protein_id="AAC67660.1"
           /db_xref="GI:3328463"
           /translacion="WLSCIFDQFIPLSSFLAVSLICMTPTALMGTILLVEROPLSESL
           SHACYPGLLIGALSIVKVPASDSLWVIFFGCLASVIGCLGISFLERKLMHKDSAL
           CLIVSFGVGVIIVSYVKDCPLNKNINAVYGOAATLTGTEAKLIIIFCLSAV
           LMMYROIISVAIFDERFAYSGGLRTRELIVLVITSLIVSGVSVGLIISAMEVA
           PPLSAROLSDRLSTLILSIFGICGALGCFYSAFTPOCVYVECKPSIILPGRV
           VFRAGVILFLCLIFSWKTGWTIRYRRKWFPLFSRDEHLKTFYLRQNTYQGMRD
           FVRSKRYOYFQDKVPRFRMEFLCKGLVSCSEHWSLTKGLARAKLVRAHRLME
           SYLSQDLFNKNVHHFAEMEHVLTDELDTLSQMLDDPDVDPHQRIPIKTRSDG
           C"
           11920. .12876
gene      /gene="ytgd"
           /note="CT070"
           11920. .12876
CDS       /gene="ytgd"
           /codon_start=1
           /transl_table=1
           /product="Integral Membrane Protein"
           /protein_id="AAC67661.1"
           /db_xref="GI:3328464"
           /translacion="

```

Query Match 54.0%; Score 883.4; DB 1; Length 14168;
 Best Local Similarity 73.3%; Pred. No. 1.6e-200;
 Matches 1160; Conservative 0; Mismatches 416; Indels 7; Gaps 2;

```

QY      1 GAATATAAAACCTATCAGATAGAAATAAAGTATTGAGAGGT-AAATATGACAAA 59
DB      7388 GATCAACAAGAAAGAACTTAACAATAAACAGCTTTTCAGAGGGAATATGACTCAA 7329
QY      60 ACCGAGAAACCTTTGGAAAATGCGCTCTCTTCTGTGGCCGATATACACTACGAG 119
DB      7328 ACCGCGGAAACCTTTTGGAAAATGCGCTCTCTTCTGTGGCCGATATACACTACG 7269
QY      120 CTAAGAAGATTCTGCCAATGTTCTAATGTTCTTGTATTACATTTAATAACGATG 179
DB      7268 CTGAAGAAGTTCTGCCAATGTTCTAATGTTCTTGTATTACATTTAATAACGATG 7209
QY      180 TTAACGATACAAAGACACTTATTTGTGGAGCTCTGTTCTGCTGACAGGCAATA 239
DB      7208 TTGAGAGATACAAAGACACTTATTTGTGGAGCTCTGTTCTGCTGACAGGCAATG 7149
QY      240 CCTTCATCAAGTTTGGCTGTTGCTGCTGCTGCTGCTATTTATCTTTATGCTTATTA 299
DB      7148 CCTTCATCAAGTTTGGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTTATTTACGCGC 7089
QY      300 AAGCTAAGTAATATTATTAAGTAAAGCGCTTATTTAAGTGGGAGCGGCTTTTAA 359
DB      7088 AAGCTAAGTAATATTATTAAGTAAAGCGCTTATTTAAGTGGGAGCGGCTTTTAA 7029
QY      360 ATTTCCTTTGCCCTGTTCCGAGCTGTAATTTATCCGCTACGGAGGATTTTACACTACA 419
DB      7028 GATTCTTCGCACTGTTCCCTGTTGAGTACACCTTGGCGGATATTTCTTACCCACAA 6969

```

QY	420	GATTGGTGCACCTTGTACACGAGCCATCCATCCGACGAGATGGTACGATCCGTGCATC	479
Db	6598	GCTTTCGCTGTACTCTACACATCCATCTCTCTCCGGATTATGGGATTCACTTGCATG	6909
QY	480	TTAAGAACTGGACATTTGCTGCATTTTATGTACTGTGAACATAATGGGAAGCCTCATG	539
Db	6908	CTACGCAACTGACATTTGCTGTGTCTTACGTACTTTCTGAACCTTTGGGAAGCCTTATG	6849
QY	540	CTATCTCTAATGTTCTGGGCAATTTGCTAATGAATTTACAATAATCCAGCAAGAGCT	599
Db	6848	CTCTCTTGGATGTCTGGGGCTTTGGCCAAAGAATTACTAAATTAAGCCAACTAAAGCT	6789
QY	600	TTTACGCTTTTTCGGTATCCGGAGCAATATTTCTTCTACGTACCTGGTGGTGCATTT	659
Db	6788	TTTACGCTCTTTTCGGAGTAGAGACTAACGTAGCTCTCTTAATTTCTGGTCCACTATT	6728
QY	660	GTTTGGGCTTCAAGTGTGAGAGCCTCCGTTTCTGAAGGTGTAGATCTTTGGGGAATTTCT	719
Db	6728	ATCTGCTCTCTAATTAATGCGGAGAGCCTTGGGGAAGGGGTTGACCACTGGGGAAGCT	666
QY	720	TTTACGCTTTTGTATGCTATGACTATTGTATCTGGACCTGTCTTATGGCCAGTTACTGG	779
Db	6668	CTCTATTTCCAAAGGCTATGTTCTGTGTCTTGGCGCTATTATTCGCTGTCTACG	6609
QY	780	TGATATCAATAAGACGTATGACCGATCCCGGTTCTATATATCCAGAAGAAATGCAAG	839
Db	6608	TGATATGAACCGCTACGCTTATCCAGTCTCAGATTTCTACAAATCTCGACAGCTCAAACT	6549
QY	840	GGGAAAAAGGTGCTAAACCTAAATGAATGAATGAAGATAGCTTCCCTATCTTATGAGA	899
Db	6548	AAGAAA-----TCTAATACCTAAGATGAGGATGGGGGAAGCTTCTTATCTGTAA	6495
QY	900	TCTCCTTATATCTTTTATTTAACTCTCTGTGGTTATTTGCTATGTATTTGGCTTAATCTTA	959
Db	6494	TCTCCTTATATGCTTCTTACGCTCTTCTAGTATTTATTTGCTACGGAATCTGCATTAACTTC	6435
QY	960	ATGCAACTGACTTGGAAAAAGTCAAGCTGAAACCTGCATATCCATATATGAATGACTATAGT	1019
Db	6434	GTTTAAATTAATCTTGGAAAAACCCACATCAAAATGCAATTTCCATATCAAAATACATCAAGC	6375
QY	1020	GAGTTCATATGGGGAACCTCTCTCTCTGGAGCGGATATCCGTAATATCATGCTATATT	1079
Db	6374	GCAATTTATGGGGAACCTCTCTTCTCTGACAGAGGTTGTATCTGTATTTGTATATGCTCTTC	6315
QY	1080	GTTTGGGTAAACGTCATTTGCTTAAATTTGGATGGTTAACTGAGGCCCTATGTCACCTCTGC	1139
Db	6314	ATCGGTGTAAACGTTATTCCTGATATTCGGGTGATTAACGAGCTTTGGTTACACCTATTC	6255
QY	1140	ATGCTTCTCTAACAGGTATACGTTTCTTCGCTCTTGTATCTTATCTTAAACCAAGCTCTT	1199
Db	6254	ATGTTTCTCTGATCAAGAGAGCTGTCTTCTTCTGCTCTGTTATCTTCAAGATCATGTCACA	6195
QY	1200	GGGCTGTGCTCTATGTTTCGGTATCAACTCTCTCATGCTAGCTGTGGTGTGGAGCTATA	1259
Db	6194	GGTTATGTTCTGCTTTTGGSAACAAACACCACTAATGCTAGCCGTTGTGTGGAGCAATT	6135
QY	1260	CAGAAATATCTTTGGAATTCACAAATACGCTCTCTTGTACTCACTAAAGAAATGGCC	1319
Db	6134	CAAAATATCTTTGCAAAATTAATTAATATGCTCTCTTGTATGCAACATAAGATAGCTCT	6075
QY	1320	TATATCCCTCTTGACCAAGGCAAAAAGTCAAGGATGAGGCTGATTTGATGTAGTTGCC	1379
Db	6074	TATATCCCAATTTGACCAAGGCAAAAAGTAAAGGAAAGCTGATTTAGAGCTTTGTGCC	6015
QY	1380	GCCGCGCTTGGAAAAATCAGAGAGAGCTTTTAATCCAAAGAGTTTCTCGTTATCTGTGGA	1439
Db	6014	GCTAGATTCGGTAAATCTGGGGGATCTTTGATTCCAACAAAGCTCTCTTGTGCTTTGGCGA	5955
QY	1440	AGTATTTGAGACTATGACCCCTTATCTTTGACAGTATCTTCTTTTCAATGATGCTATTGG	1499
Db	5954	AGATATGGGCGCTATGACTCATCTTCTAGAGCTTTGTCTCTTTGCAATTTATCATGTGTGG	5895
QY	1500	TTGGTTTCTGCAACTAAGTTAAACAACATATTTCTTAAGGCGAGTCTGCTTTAAAGAACAA	1559

Db	5894	TTGAACTCTGCAACTTAACCTTAACCAACTGTTCTTGCGTGCATCCGCTGCTTAAGAACAG	5835
OY	1560	GAAGTGCTCAAGAAGATTACG 1582	
Db	5834	GAATTAGCAGAAGCTGCAGCAGC 5812	
RESULT 5			
LOCUS	TAJ10586		
DEFINITION	Chlamydia trachomatis npt1 gene.		
ACCESSION	AJ010586		
VERSION	AJ010586.1 GI:4538859		
KEYWORDS	Npt1 gene; nucleoside triphosphate transport protein 1.		
SOURCE	Chlamydia trachomatis.		
ORGANISM	Chlamydia trachomatis		
REFERENCE	1. (bases 1 to 1587)		
AUTHORS	Tjaden, J., Winkler, H.H., Schwoppe, C., Van Der Laan, M., Mohlmann, T. and Neuhaus, H.E.		
TITLE	Two nucleotide transport proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy		
JOURNAL	Bacteriol. 181 (4), 1196-1202 (1999)		
MEDLINE	99138740		
PUBMED	9973346		
REFERENCE	2. (bases 1 to 1587)		
AUTHORS	Neuhaus, E.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-AUG-1998) Neuhaus E., University of Osnabrueck, Biology / Plant Physiology, Barbarastr.11, D-49069 Osnabrueck, FRG		
FEATURES			
source	1. 1587		
gene	/organism="Chlamydia trachomatis"		
CDS	/db_xref="taxon:813"		
	1..1587		
	/gene="npt1"		
	1..1587		
	/gene="npt1"		
	/function="ATP/ADP transport protein"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=1		
	/product="nucleoside triphosphate transport protein 1"		
	/protein_id="CA839534.1"		
	/db_xref="GI:4538860"		
	/translation="MPTGAEKPEKTLKSLFIMHMLKYLPMFLMEFCSFNYTLLRDKDITLITAPGSGARLPIETIKLMLVPSAVVHMLYAKSLNLKOLFFAVLSPPVFEALFPVAVIPCRHILHPAEADTLQSLIPSEMGFIAMLRMTAVAVYVLSLGMGVSVMLSLMEGFAVNEITRKISEAKREYALDVGANVALISGPAITMSKLRASLGEVDPMGSLYFLMAMFLFSCAIIIAACYMMNRVYLDPFRYNPAELKAKSKPKMSGSEFSYLLRSYMLLALVLICVIGICINLVENWYSOLKMOFPNDVSAFMGNSFWTGVVSVPMFLFIGNVJIRFGMLGALVTPIMVLTGAVFPALVIFRHATGLVAALGTPPLMLAVVATVNIILSKSTKRYALFDATREKLATIPDQEKVYKKAIDVYARFSGSGSLDQGLLVCGSIGAMPFLVALFLAIIMWLTISATLKLNLKFLAASAKKDELAEAAA		
	AERKASSAARESAALIEGVS"		
BASE COUNT	384 a 357 c 330 g 516 t		
ORIGIN			
Query Match	52.9%	Score 865.2;	DB 1; Length 1587;
Best Local Similarity	73.3%	Pre. No. 4.3e-196;	
Matches 1123; Conservative	0;	Mismatches 403;	Indels 6; Gaps 1;
OY	51	ATGACAAAACCGAAGAAACCTTTGGAAATTTGGCGCTCTTCTTGCGCGATACAT 110	
Db	1	ATGACTCAACCGGGGAAAAACCTTTGGAAATTTGGCGCTCTTCTTGCGCGATACAC 60	
OY	111	ACTCAGCAGCTAAAGAAAGTTCTGCCAATGTTCCTAATGTTCTCTGATTACATTTAAC 170	
Db	61	ATGCATAGCTGAGGAAGAAAGTTCTGCCAATGTTCCTAATGTTCTCTGATTACATTCAT 120	
OY	171	TATACAGGTGTTACGCGATACAAAGACCTCTTATTTGTGGAGAGCTCGTTCCTGGTGCA 230	

[illegible]

gene
CDS
/transl_table=1
/product="RNA polymerase sigma factor, sigma-70 family"
/protein_id="AAF39194.1"
/db_xref="GI:7190373"
/translation="MKTODLADTQWLQVSTKEIHHRDFLEAYLPVLYKNNVAHRLASGM
PSHVIEDLVASGVEGLVRAVEREDPEKSKRFESYALPTIKAAIIDLRKODWPRSV
YORANRLADAMDALROTIGKEPTDGLCEYINISQOELSHWSSRPALVSLINDDES
CODEDEGLAEERLADPERAENGVDYIRKKEAISLTETALSLDEKEREYMAIYYDDL
VLKEGTIGVSESRVSOIHSKALLKRLSSL"
2869. .4107
/gene="TC0332"
2869. .4107
/gene="TC0332"
/note="similar to SP:P00952; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="tyrosyl-tRNA synthetase"
/protein_id="AAF39195.1"
/db_xref="GI:7190374"
/translation="MQQLIDSLQKRGILDNSSAGLESITAPVASYLGFDPAPAPLHIG
HWIGICFLRLRSAYGITPILAVGATGTGIGPSSKRSVERSLDDEQVLDNSKLEVAL
ANYLPDIRVNNADMGLSLMVDLRLDGKRYRGLSMKADVVKORYSEGISYTER
SYLLDSYDFALHFKHGVVLCQSGSDQMGNTSGIDYIRKGLGQAFGLYPLTDS
KGGIKGTESTQWLDPELTPSYELFOYFLRLSPOEIPKIRMLTLDDDEVLADR
LENDQAVKRYVAIVIVKDVYAGAGLAAQALATPFSFANKCKNTSESRLAVOSGVG
INVARADYIGKRMVDVYVQGFSGSSKGEARLIDQKGLYVQPEPLIDQSVLDGTILC
FDRIYLSQGRKKKQVIDLN"
4117. .5556
/gene="TC0333"
4117. .5556
/gene="TC0333"
/note="similar to SP:P00349 PID:1193; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="6-phosphogluconate dehydrogenase,
decarboxylating"
/protein_id="AAF39196.1"
/db_xref="GI:7190375"
/translation="MAPADIGLGLAVGKNVLMNTDGHFAVSVNRSEKTEEFK
EHGENISLQGTALIEFVOSLTKRPRIIMIKAPVDEMISLLPFLERDILLIDG
NSYLDSEERYIDLKKGILFLVGMVGSGEGARKGSIIMPGENIEMAPYAPIFOI
AAVDGQPCSMITGGAGHVKVANGIEGDIQICETMYEILKSLRDISLBOIGM
FEFWNOTDINSYLSMASAVLTAKDENGVAASITIDVACQKGTGRVADAKAGP
MSLITESVLARYLSAKMEVRQARREPPVASLTPQSEASVLIEDARELVIKATIS
YAOGEFLIKQISEERNWDLNIGELALWRGCTIIQSAFLDKIHOGFESCPDASHLMLO
DYKFNVLINSETGFRRAIILHAGVAGVAPICLASALAFYDGYRTENSPLFVQGRDYF
GAGYERODRPRGEFYHTDMLGSKNASRM"
complement(5624. .7432)
/gene="TC0334"
complement(5624. .7432)
/gene="TC0334"
/note="similar to GB:D17650 SP:P37949 PID:1122398
PID:436036 PID:1303804; identified by sequence similarity;
putative"
/codon_start=1

gene
CDS
/transl_table=1
/product="GTP-binding protein LepA"
/protein_id="AAF39197.1"
/db_xref="GI:7190376"
/translation="MKPKIENINFSIAIHDIHCKSTIADRLLESTTIFOREMREO
LLDSMDLEREGITIKAPVMTTEYBESETELLIDTPGVNDSYEVSNLAACEGA
LLIIDAQGVQASLANVYLALEPDELITPVLNITDIPAAQPEAIKQIEFFGLDTS
NAIACSATGGGIEPEILESIRLPPRPSQSETEKALIFDSHAPVYGIMVYRVIS
GEIKKDRITPMSTKSGSEFVIGICAFLEATLIEGSLRAGOVGYFIANKKVDKI
GDVTATKHPAKPELEGEKEIKPVFGIYIDSSDPTLDKALGRLOLNDALITIO
ESSHIGFGRFGEGFLIHLEIFERISREPDILITAPSVIKYKILKNGKTLFON
PTAYPDRLAIHMEEPVHVNIITPQETLSNMSLCMDKRGICLKTMLDQRLVLS
ELPLNEIVSDNDLKSVTKVGSFDRGLGDKGALIKLEILINDETVAQSCVLR
DKAASKRSICEKLVYDIPOLFKPIQOALNKKITARETIRALAKNVTAQCYGDIT
RKRLMKQKKGKKRMEKFCVSIPTAFVLEKME"
complement(7617. .9206)
/gene="TC0335"
complement(7617. .9206)
/gene="TC0335"
/note="similar to GB:M28816 SP:P19568 PID:152470
GB:AJ235269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ADP, ATP carrier protein"
/protein_id="AAF39198.1"
/db_xref="GI:7190377"
/translation="MTQAEKPEFGLRSFLPIMHHELKVLPFLMFCISFNVTIL
RDKDTLIVTAPSGSAEAIPIKMLVPSVAVMLIYAKLSNITSQALFEVLSPF
VVFPAFLPVIYPRRIITHPDADFDLQALIPSGEIGFIMLRMMPFAFVYSELNG
SVMLSLMPGFANETITKISEKRRYALFGCANVALLISGPATYWSKRLASIGEVD
PWGVTLYFLAMFSCAIIAACYWMNNRVLVDPDRYNAELKAKSKRPMMSGES
SYLLRSPYMLLALVLVICYGVCINLVYVTRKSQMKPMPNEYSAPMGFSFMTGV
SVFWMFLTGNVIRFCMLGALVTPMVAVTGCAIFPALYIFRDHATGLVAALGTPPL
MLAVGCAVONIISKSTKVALFEDATKEMAYIPLDOEGKVGKAIDVVAARFCKSGS
LIOOGLVVGGSISAMPPLAVALFALINWLSATKLNLPLAASAKKQDELAEATA
AAKEBASPAKKEVPAIEGVS"
complement(9533. .10029)
/gene="TC0336"
complement(9533. .10029)
/gene="TC0336"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF39199.1"
/db_xref="GI:7190378"
/translation="MATROTTIQEIEIQLTKAISKVGSKENDCLRYLPGPCTYH
HFTLKMKHSAPEOLLMLKTFVYIDSWTPRSINKPRAPRGSKRKRDVNFVFTDIER
VLELARQVGKDLIARFSPKPLPSLKRRELRISIRNIVSTELMNVVVALSQPTE"
complement(10169. .10270)
/gene="TC0337"
complement(10169. .10270)
/gene="TC0337"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF39200.1"
/db_xref="GI:7190379"
/translation="MOTNLNNLKHPIFCPNOKISKMLRKQKPL"
10658. .11638
/gene="TC0338"
10658. .11638
/gene="TC0338"
/note="similar to PID:790546 PID:1777933 GB:AE000520;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ABC transporter, periplasmic substrate-binding
protein, putative"
/protein_id="AAF39201.1"


```

/db_xref="GI:7190380"
/translation="MFELHVRKRYKHVIGLGLIACGEVIVSSCGSRGKRSIDERIHIL
SMNRIVDCVRIQRYKNIIVLIDGSDIPSYEMKVEDERMIISQIIFPNGLGLEH
SASLRKHEGNSKYLDIGARILDKNCFPLISEDFPDHITMDGVMISYKEMASVL
VOQIPEYAEEROKNAEQLISEMEDIDRMAVSLATPEKKNYLVTEGHAFSYFTRRYL
SSDERESGNWKLRCMKSPESGLSPDAOISIRIDIMRVEVICANDGVFLEEDTLNODAL

Query Match      52.7% Score 862.6 DB 1: Length 23773:
Best Local Similarity 72.5% Pred. No.1.4e-195;
Matches 1147; Conservative 0; Mismatches 429; Indels 7; Gaps 2:

QY 6 AAAAAATATCAGATAGAAAATAAAGTATTTTCAGAGGT-AAATATGCAAAAACGA 64
Db 9252 ACAAGAAAACCTTAAACAATAAAGCTTTTCAGAGGTGAADATGATCAAAACGCC 9193
QY 65 AGAAAAACCTTTGGAAAAATTCGCTCTTTCTTGCGCGATACACTACGAGCTAA 124
Db 9192 GGAAGAACCTTTTGGAAAAATTCGCTCTTTCTTGCGCGATACACTACGAGCTAA 9133
QY 125 GAAAGTTCGCCAATGTTCTTAATGTTCTTCTGATTTACATTAATACATGAGTTCAG 184
Db 9132 GAAAGTTCGCCAATGTTCTTAATGTTCTTCTGATTTACATTAATACATGAGTTCAG 9073
QY 185 CGATACAAAAGACACTTATTTGTGGAGCTCTGTTCTGTTGCGAGAGCAATACCTT 244
Db 9072 AGATACAAAAGACACTTATGTCACAGACAGGCTCTGAGCAGAGCAATTCCTT 9013
QY 245 CATCAAGTTTGGCTTTGTTGTCCTCGCTGCTATTTATCTTATTTATTTATGCAAGCT 304
Db 9012 TATTTAGTGTGGCTTTGTTGTCCTCGCTGCTATTTATCTTATTTATGCAAGCT 8953
QY 305 AAGTATATTTTAAAGTAAAGCAGGCTTATTTATGAGGAGGAGGCCCTTTTATTTT 364
Db 8952 AAGCAATATTTTGAAGCAACAGGCTTATTTATGAGGAGGCTCTCTCCATTTGTTGTT 8893
QY 365 CTTTGCCCTGTTCCGACTGTATTTATTCGCTACGCGATGTTTACATCTACAGAAAT 424
Db 8892 TTTTCACAGTTCCTTTGTTGATTTACCATATGCTATATGCTTACCCAAAGACTT 8833
QY 425 TGCTACCGCTTTACAGGCGCATCTCTACCTCAGAGATTGCTAGAGCTGTTGCCATCTTAA 484
Db 8833 CGCTATATACACACAGGCGCATCTCTCGGTTCTTGAAGCTTATGCGCATGCTAG 8773
QY 485 AACTGACATTTGCGATTTTATGTACTGCTGACATGAGGAGGAGGTCATGCTATC 544
Db 8772 CAACGTGACTTTTGCAGCGCTTTTATGTTCTTCCGACTTTGGGAGAGTATGCTTTC 8713
QY 545 TCTAATGTTCTGGGATTTGCTAATGAATAAATCAAAAATCCAGAAAGCGTTTCTA 604
Db 8712 GCTAATGTTCTGGGATTTGCAAAATGAATACTAAATTTAGCGAAAGCTTATCTTA 8653
QY 605 CGCTCTTTTGGTATCGAGCTAATATTTCTTTACTAGCTTGTGCTGCAATTTGTTG 664
Db 8652 CGCTCTTTTGGTATCGAGCTAATATTTCTTTAATTTCCGCTGCAATTTGTTG 8593
QY 665 GGCTTCAAGTGTGAGAGCTTCGTTTCTGAAGTGTAGATCCTTGAGGATTTCTTTAG 724
Db 8592 GTCTTCTAATATACCGGAGCTTGAAGTGAAGGCGAGACCATGAGGAGTACTCTTTA 8533
QY 725 TCTTTTGAATGCTATGACTATTTGATCTGAGCTTTGCTTATAGCCAGTATCTGAGAT 784
Db 8532 CTCTTAATAGCTATGCTGCTGTTCTGTCGCAATTAATCGCTGCTGCTACTGAGAT 8473
QY 785 CATATGAAGCTATGACCGATCTGCTCTTCTAATATCCAGAGAAATGCAAAAGGGGAA 844
Db 8472 GATATGTTAGCTTCTTACGATCTAGATTTCTAATACCAGAGCTAAAA-----GC 8419
QY 845 AAAAGTGTAAACCTAAAAATGAATGAATGAGTCTCTCTATCTGATGATATCTCC 904
Db 8418 TAAAAATATTAACCTAAGATGAGTGGGAGAGCTTCTCTTATCTTTAAGATCCCC 8359
QY 905 TTATATCTTTTATTAACCTCTGCTGTTATTCGCTATTTGCTATTAACCTAATCGA 964

```

```

Db 8358 TTACATGCTTCTTACGCTTCTGATTTATTTGCTACGAGTTTGCATCAACCTTGTGA 8299
QY 965 ACTGACTGGAAAGCTCAAGCTCAAACTGCAATATCTTAATATGATGACTATAGAGAT 1024
Db 8298 ACTAATCTGGAAGACCCACCTCAAAATGCAATTCCTTAATCCAAATGATATACAGGCAAT 8239
QY 1025 CATGGGGAACCTCTCTCTCTGAGTGGCGTACTATCCGACTATATCATCTTATTTGG 1084
Db 8238 CATGGGGAACATTTCTCTTCTTGACAGGAGGTTTATCCGATTCGTTATGCTTATTTGG 8179
QY 1085 TGTATAGCTATTCGTAAATTTTGATGTTTAACTGAGAGCCCTAGTCACTCTGTATGAT 1144
Db 8178 TGTATATGTTATTCGTCGTTTGGCTGTTTACAGGCGCTTATGATACCTGTCATGAT 8119
QY 1145 TCTCTAAGAGTATGCTTTTCTTGCTCTTGTATCTTTAGAAACCAAGCTTCTGGCT 1204
Db 8118 CCTGTTAAGAGAGGAGTCTTCTTGCTCTCTTATCTTCAAGATCATGCAAGAGAT 8059
QY 1205 GGTGCTATGTTGCGTACACTCTCTCATGCTAGCTGCTGTTGCGAGCTATACAGA 1264
Db 8058 GGTGCTAGCTTTGGGACACACTCTTATATGCTAGCGGTTGTTGGAGCCGTCAGAA 7999
QY 1265 TATCTTTGAAATCCACAAATACGCTCTCTTGAATCACTAATAAGAAATGCGTATAT 1324
Db 7998 TATCTTTCCAAATGACTAATATATGCTCTTTGATGCACTAAGAGATGCTTACAT 7939
QY 1325 CCTCTTGACCAAGCAAAAGCTCAAGATGAGCTGCTATGATATAGTTGCCGCCG 1384
Db 7938 CCTATTTGATCAAGAGCAAAAGTAAAGAAACCGCTATGACGTTGTCGCTAG 7879
QY 1385 CTTCGAAATACAGAGAGGCTTTAATCCAAAGGTTTGCCTGTTATCTGTTGGAATAT 1444
Db 7878 ATTGATTAATCTGGGGGATCTTTTATCCAAAGAGACTCTAGTTGTTGGGAAAGAT 7819
QY 1445 TGGACCTATGACCCCTTATCTTGACAGTATCTCTTTTCACTATGCTATTTGGTGGT 1504
Db 7818 CAGTGCAGTATCCATTCCTGAGCGTACCTCTTTGCAATTAATATGTTGGTTAAC 7759
QY 1505 TTCTGCACTAAGTTTAACAAACTATCTTTAGCGAGTCTGCTTTAAGAAACAGAGAT 1564
Db 7758 ATCTGCAACCAAACTTAATAATTTGTTTACCTGCTTCAAGCTGCTAAGAAACAGAGCT 7699
QY 1565 GGTCAAGAAAGATTCAGCTCTG 1587
Db 7698 GGCAGAGGCCACAGCAGCTG 7676

```

```

RESULT 7
AX098423 1996 bp mRNA linear PAT 02-APR-2001
LOCUS
DEFINITION Sequence 2 from Patent WO0120009.
ACCESSION AX098423
VERSION AX098423.1 GI:13537715
KEYWORDS
SOURCE
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 1896)
AUTHORS Moehlmann,T. and Tjaden,J.
Plants having altered amino acid contents and method for the
production thereof
Patent: WO 0120009-A 2 22-MAR-2001;
BASEF AKTINGESELISCHAFT (DE)
FEATURES
location/Qualifiers
source
1..1896
/organism="Solanum tuberosum"
/db_xref="taxon:4113"
BASE COUNT 473 a 369 c 445 g 609 t
ORIGIN

```

2.

999

```

/translation="MEGVLOTBGLSLPSKPKIKAFPLPQGLNRNLSLSKPNP
LNGVLSLSNGFOKVOGPDTPQLPFGOKRCEPCIEKAEBAAGAAAGDGLPEKEDPK
FMGIELATLTKRIPIGAMFECILFENYTLIRORXDLVLYAKSSSEIIPLEKTNL
MAIGFMLYTKLANLSKALPYTYVILPIAFGAEVLYPLSLFHEBTADKLN
TLGPRFLPILIRLWISFCLFYMAELMWSVSVLFWEFAOITTVDEAKFPELFG
LGNALJESGRIVYFSSLRSLPGVGNALISLKGMSIYVMGGAACFEPYWNR
NVALPJRKKKKVKNMTMESLAKFLVSKIRLADTLVAJISINLVEVWRSKLK
AOFSPNEXSSFMGPDSTAGIATFTMMLSQWIDFKYGMGAATITPVILLTGVEF
PSLLFAPLAPTLAKFGMTPELLAAVYUAMONITSKSKYSLPDPCKEMAITPDED
TRVKKAIDIVCNPLKSGGALIOQFMILIFESGLASSPVLGVLVIVLWLAAK
SLDGGFTOLROBEDLEKEMERASLIKIPVVSOMENGNPLSSESLNPAGDSTMASSE
PSSPSRL"
BASE COUNT      590 a      443 c      518 g      749 t
ORIGIN
Query Match      23.2%; Score 380.2; DB 8; Length 2300;
Best Local Similarity 55.1%; Pred. No. 3.1e-80;
Matches 804; Conservative 0; Mismatches 628; Indels 27; Gaps 2;

74 TTTTGGAAAATTCGCTCTTCTTGTGGCGGATACATACAGCAGCTAAGAAAGTCT 133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
411 TGTGAAAAGAGACAACTAAGTTTATGGGATTTGAACTTGACCCCTTAAGAAAATTAT 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 134 GCCAATGTTCTTAATGTTCTTGTATTAATTAATACGCTGTACGCGATACAA 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 ACCACTTGGGGCGATGTTCTTTGTATTCGTATTAATTAACAATCCTTAGGATAC 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 194 AGACACTCTTATTTGGAGCTCCTGCTGTGTCGACAGGCAATACCTTTCAATCA 253
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 GGATGTGTGGTGTACAGCAAAAGGATCCAGTGTGATTAATCCCTTTCTTGA 590
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 254 TTGGCTTTGTGTCCTGCTATTAATCTTAATTAATTAATTAATTAATTAATTA 313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 TTGGGTAATTTGCTATGCTATGATTAATGATTAATGATTAATGATTAATGAT 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 314 TTTAAGTAAGCAAGCTTTAATTAATTAATTAATTAATTAATTAATTAATTA 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 GTTGCAAGAGAGCTCTTTTATTAATGATTAATGATTAATGATTAATGATTA 710
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 374 GTTCCGAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 GTTGGTTTTGTTTGTATTAATTAATTAATTAATTAATTAATTAATTAAT 770
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 434 TTTACAGGCAATCTACCTACAGATTTGTCAGAGATTCGATTAAGAAATCG 493
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 GCTTCAATACCTTGCTGCAAGATTTCTGGAACCAATGCTTAATCTGAGAT 830
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 494 ATTGCTGATTTTATGCTACTGCTGCAATGAGGGAAGCGCTAATCTTAATG 553
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 TTTCTGCTGTTGCTATGCTAGCTGAGCTTTGGGAAGTGTGTTCACTACTCT 590
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 554 CTGGGATTTGCTAATGAATAATTAATTAATTAATTAATTAATTAATTAAT 613
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 891 TTGGGATTTGCTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 950
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 614 CGGTATCGAGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 673
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 951 TGGACTTGGAGCAATGTTGCTTTATTTCTTGCTGACAGCTGAAGTACTTTCT 1010
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 674 GTTGAAGCTTCCGTTTCTGAAGCTAGATCTTGGGGAATTTCTTAAGCTTTG 733
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1011 CTTGAGAAGCTTTTGAAGCTGAGTGTGTTGGGCAATCTCCCTGAAGAAATG 1070
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 734 GGTATGACTATTTGATCTGAGCTTGTCTTAATGCGAGTACTGATCAATAGAA 793
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1071 GAGTATGTTGTGATGATGATGGGCAATCTTCTTAACTGCTGGGTGAATGAA 1130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 794 CGATATGACGATCTCGCTTATTAATCAAGAAATGCAAAAGGGAAGAAAGTGC 853
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1131 TGTGCTCTCCCACTGTAAGCAA-----GAGGAAGAAAGG 1166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 854 TAAACCTAAATGAATATGAAGATAGCTTCTCTATCTTGAATAGATCTCTATAT 913
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1167 AAAACCTAACATGACCAACATGAGAGCTTGAAAGTTCTTGCTCTTCAAAATATAT 1226
QY 914 TTTATTAACCTCTTGTTGTTATTTGCTTAATGTTGATTTGATTAATTAATGAAGT 973
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1227 GATCTTGGCACATTTGTTGTATGATATGATTAATTAATTAATTAATTAAT 1286
QY 974 GAAAGCTGACCTAAAGCTGCAATATATGATTAATTAATTAATTAATTAATTA 1033
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1287 GAAGTAAAGCTTAAGAGCTAGTTCCTCAAGCCCAATGATTAATTAATTAAT 1346
QY 1034 CTTCCTCTCTGAGCTGCGGTAGTATTCCTTAATTAATTAATTAATTAATTA 1093
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1347 CTTCCTCACTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1403
QY 1094 CATTCGTAATTTGATGATGATTAATTAATTAATTAATTAATTAATTAATTA 1153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1404 TTTTCGACAAAGTATGAGGTGGGAGACACAGCAAGATTAACCTTAACATCT 1463
QY 1154 AGGTATCGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1464 CGAGTGTGTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1523
QY 1214 GTTGGTACAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1524 GTTGGAAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1583
QY 1274 GAAATCACAATTAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1584 TAAGAGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1643
QY 1334 CCAAGAGCAAAAGTGAAGCTGATTAATTAATTAATTAATTAATTAATTAAT 1393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1644 TGAAGCACCAAGATTAAGAGGAGGAGCAATGATGTTGTGCAATCACTGG 1703
QY 1394 ATCAGAGAGAGCTTAATCAACAAAGTGTGCTGTTATCTGTGAAGATTAAT 1453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1704 GTCTGGAGAGCTTGTGATCAACAGTTCATGATTTGATCTTTGTTGTTG 1763
QY 1454 GACCCCTTATCTTGAAGTATTTCTTTTCAATCAATGATTAATTAATTAAT 1513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1764 GACACCCCTCTGAGGAGGCTCTCTGATTAATTAATTAATTAATTAATTA 1823
QY 1514 TAAGTAAACAACTATTC 1532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1824 GTCTTTGATGACAGTTTC 1842

```

RESULT 9
 GSU251356
 LOCUS
 DEFINITION
 accession
 version
 keywords
 source
 organism
 reference
 authors
 title
 journal
 reference
 authors
 title
 journal
 source

GSU251356 2126 bp mRNA linear PIN 04-JAN-2002
 Galidieria sulphuraria chloroplast mRNA for plastidic ATP/ADP
 transporter (atp1 gene).
 AJ251356
 version 1.1
 keywords atp1 gene; plastidic ATP/ADP transporter.
 source Galidieria sulphuraria.
 organism Galidieria sulphuraria
 Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;
 Porphyridiaceae; Galidieria.
 reference 1
 authors Stamme, C. and Neuhaus, E. H.
 title Analysis of the plastidic ATP/ADP transporter from the red algae
 Galidieria sulphuraria
 journal Unpublished
 reference 2 (bases 1 to 2126)
 authors Neuhaus, E. H.
 title Direct Submission
 journal Submitted (26-Nov-1999) Neuhaus E. H., Plant Physiology, University
 of Osnabrueck, Barbarastr. 11, 49069 Osnabrueck, GERMANY
 location/Qualifiers
 source 1. 2126
 /organism="Galidieria sulphuraria"
 /organelle="plastid"

gene
CDS

1. 2018
/gene="aatp1"
/gene="aatp1"
<1. 2018
/gene="aatp1"
/codon_start=3
/transl_table=11
/product="plastidic ATP/ADP transporter"
/protein_id="CAC80882.1"
/db_xref="GI:18073248"
/translation="EMVQPSFVFINLVHSSHMWKTRQSVSRKNNVAHFCPKETSRF
SOAGRTCECHKSGSETFYRNEAASNNELSTARLESEHERIALSSYPVAPPQIH
SNSSTLSLATATVATLGQWFKLVSRKADVDLLSPQFLQVSSAGTSGSAGVAG
DGIALDPKRNELKKILPLGVMPFMLENATVILLRPTKVILVTTGTGAKI IPELTVA
LPGVATPIAASKLSNMMDRETLFVYCVPIILPISAFVLPYELRHALHPYAVDMY
AGYLPASTFAFLPGITIRNNTFLFTFLALMGSVVALFMPGFANEVSLVDKAKTYP
FGLVAANVALFDSGOIVRFVSDIRKLLPPGVQDVGSLKYLMTMIGVAGSVIILMLVY
QKNVLTDPNCVDMSKRNKTKQKTLGKESAVYLAKSAYIRNLALVYAGMSINIVE
VSMKSLKEAFDPNDSYSTGMGWESTGTSVTLIMMLLGR IIPKFGGFALATPTVY
LGITFGFCLTFLFSROLHPVVMGIGTFLMLAVLYAGNAONILSSKSKSLFDCKEK
AYITLDAEOKSKGAALDIVIGNPMKSGGSGFLIOOGLIFAVGSLAASPYYLAFILFV
GMNLTAAASLSDKQFOEAKRKSVEVAVMAKTIQDPIEAAKVAKESKQ"

BASE COUNT
ORIGIN

567 a 388 c 467 g 704 t

Query Match	22.9%	Score 374.8	DB 8	Length 2126
Best Local Similarity	55.6%	Pred. No. 6e-79		
Matches 790; Conservative	0	Mismatches 617;	Indels 15;	Gaps 3;

OY	115	ACGAGCTAAACAAGTTCTGGCAGTATGTCMAATAGTCTCTGTATACCTTAATACTA	174
Db	514	ACGAGCTGAAAAAATCTCCCTCCGTTGGGTGTCAATGTTTATATGATTTCTATTAATTA	573
OY	175	CGGTGTTACGGGATACAAAACACCTCTATTGTGGAGGCTCCGTGTTGGTGCAGAGG	234
Db	574	CTATCTCGAGAGATACAAAAGACGTTCTCGTAGAGCAACTACAGGA-----CGAGAA	627
OY	235	CAATACCTTCATCAAGTTTGGCTGTGTGCCCTGTGCTATTACTTATGCTATT	294
Db	628	TCATACCGTTCTTAAACATATGCCAATCTCCGTGAGGAGCTTATTTACTATATGGCT	687
OY	295	ATGCAAGCTAAAGTAATATTTTAAGTAAGCAGGCTTATTTATGCAAGGGAACGCGCT	354
Db	688	ATTCAAAGTTATGCATATATGTTTATGCGGAACACTATTTATGTTTATATGCTTCCT	747
OY	355	TTTTAATTTCTTTGGCCCTGTGCCGACTAATTAATCCGCTACGCGATGTTTACATC	414
Db	748	TTATATATATTTTTCATATTCATTGCTCTTTGTGATACCGCTTGAGACATGCCTTATC	807
OY	415	CTACAGATTTTGTGACCGTTTACAGCCACACTACCTCCAGATTTGCTAGAGCTGTTG	474
Db	808	CTTATGCCCTTTGTGAGATTGGATACGTGATATCTTCCGCTTCCTTTTGTGCTCTTGTG	867
OY	475	CCATCTTAAAGAACTGGACATTTTGCTGCATTTTATGTACTTGTGTCGACATAGGGGAAGC	534
Db	868	GAATCATATCCGAATTTGGACCTTTTGCTTTCTTACACCCCTGGCGAATATATGGGATCTG	927
OY	535	TCATGCTATCTCTAATGTTCTGGGGATTTGCTAATGAATAATTAACAAAATCCAGAGCA	594
Db	928	TTGTGTATCTCTCTCTTTTGGGGTTTTCATATGAAGTAACCTTCAGTAGAGAGCA	987
OY	595	AGCCTTCTACGCTCTTTTGGTATCGGAGCTATATTTCTTACAGGCTCTGCGCTG	654
Db	988	AAAGTATATATCTCTATTTTGGCTGTGTTGCGAATGTGCTTATTTTTCAGGGCACT	104
OY	655	CAATGTTTGGGCTTCAAAAGTTGAGAGCTTCGTTTCTGAAGGTGTAGATCCTTGGGAA	714
Db	1048	ATGTACACTTTGTTTCAGATATTCGTGGAAGTTTACACACTGCTGTGGANGTGGGGTG	1107
OY	715	TTTTCTTACGCTTTTATAGGTATATGCTATTTATCTGAGACTGTCTTATATGGCCAGTT	774
Db	1108	TTTTATTGAATATTTTATATACAAATAGAGAGCTGCTGGTTCAGTCAATCATCTTTGCA	1167

QY	775	ACTGGTGATCATATAGAAGCATTTAGCCGATTCCTCGCTTCTATATCCAGAAGAAATGC	834
Db	1168	TGCTATATATTTAGAAAAATGTCGACCGAATCC-----CAATTCGTGTGATATGAGTA	1221
QY	835	AAAAGGCGAAAAAGGTGCTAAACCTTAATGAATATGAACATAGCTTCCTATCTTG	894
Db	1222	AAAAGAAACACACTAAACAAAAGACCACTGGCGCTTAAAGAGAGTGTCTTTATCTGG	1281
QY	895	ATAGATATGCTTATATTTCTTTTAACTCTCTGGTTATTCCTATGTAATTTGGATTGA	954
Db	1282	CCAGTCACACTTATATTCGAACTTGCGTTTGTAATCATGTGGATGTGATGTGATTGA	1341
QY	955	ACTTATATGCAATGACTTTGGAAAAAGTCAGCTGAACCTGAATATCTTAATATGATGACT	1014
Db	1342	ATATTTGTGAATATCATCATGAGAACCAATTTAAAGAAAGCCCTTCGTATCTTAATCTT	1401
QY	1015	ATAGTAGTTCATAGGGAGAACTCTCTCTTGACAGTCGAGCTATATCCGACTATATCATGC	1074
Db	1402	ATTCAAACCTTATATGGGTGGTTCTCACTGTATCTATGTTCTGGTAAAC--ACTTATCATGA	1458
QY	1075	TATTTGTGGTGTAAACGTCATTCGTAAATTTGGATGTTAACTGAGAGCCCTAGTCATC	1134
Db	1459	TGCTTCTTGTCGATTTATTTTCCGAAAGTTGGTTGGGATTTGGAGCTCTTATATATC	1518
QY	1135	CTGTATATGTTCTCTCAACAGATATCTTTCTTGCTCTTGTATCTTTAAGAAACCAAG	1194
Db	1519	CAACTGTTTGGGTATTTACTGGCTTTTAAATTCCTTCTTTAACTCTCTTATGTCGCACT	1578
QY	1195	CTTCTGGGCTGGTCCTATCTTCGGAGACAACTCTCTCATGTAGCTGGTGTTCGGAG	1254
Db	1579	TGCATTCACACTGCTTGGAAATCTTTGGAGACACACTTTTATGCTTTGCAGTGTCTTTGGTG	1638
QY	1255	CTATACAGAAATATTCCTTTCGAAATCCACAAATAGCCTCTCTTTCACCTAACAAAGAA	1314
Db	1639	CTGCACAAATATTCCTTCTTAATGAGACAAATATTCGTTGTGATCTTGCAGAGAA	1698
QY	1315	TGGCTATATCCCTCTTGACCAAGACAAAGTAAGAGTAAAGCTGCTATGATGAG	1374
Db	1699	TGGCTTATATTCCTATGTCAGCTGACCAAGAAATCTTAAAGAAAGCTGTCAATCGATTA	1758
QY	1375	TTCGCGCCGCTGGGAAATATGAGAGGAGCTTATATCCAAAGGTTTGCCTGATCT	1434
Db	1759	TTGGCAATTCATGAGGCAAGTCTGGTGGCTCTTATTCAAACAGGCTCTTATTTTGGAG	1818
QY	1435	GTCGAGATATTTGAGCTATGACCCCTTATCTTCGACGTGATTCCTTTTCATCATTTCTGA	1494
Db	1819	TGGGTTCCTTACACGCTTCACACACCTCTATCTGCTTTATATGTTTGTATTTGGTA	1878
QY	1495	TTTGGTGTGTTCTGCACACTAAGTTAAACAACTATCTTAG	1536
Db	1879	TGTGATAGCAGCAGCTGTTCTTTAGACCAACAGTTTCAAG	1920
RESULT 10			
LOCUS	AY098893	2122 bp	mRNA
DEFINITION	Citrus hybrid cultivar plastidic ATP/ADP transporter mRNA, partial		
ACCESSION	AY098893		
VERSION	AY098893.1	GI:21063924	
KEYWORDS			
SOURCE	Citrus hybrid cultivar.		
ORGANISM	Plastid Citrus hybrid cultivar		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.		
AUTHORS	Li, C., Weiss, D. and Goldschmidt, E. E.		
TITLE	Citrus mRNA for plastidic ATP/ADP transporter		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2122)		
AUTHORS	Li, C., Weiss, D. and Goldschmidt, E. E.		
TITLE	Direct Submission		

AUTHORS	
Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.	
TITLE	
Arabidopsis cDNA clones	
REFERENCE	
2. (bases 1 to 2110)	
JOURNAL	
Direct Submission Submitted (05-OCT-2001) Salk Institute Genomic Analysis Laboratory (SIGAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.	
COMMENT	
The Salk, Stanford, PEGC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.	
FEATURES	
Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as pls.	
source	
Location/Qualifiers	
1. 2110	
/organism="Arabidopsis thaliana"	
/db_xref="taxon:3702"	
/chromosome="1"	
/clone="RAFL09-51-C18(R14701)"	
/note="ecotype: Columbia"	
1. 55	
56. 1930	
/note="adenine nucleotide translocase"	
/codon_start=1	
/product="Atlg80300/F5f6_5"	
/protein_id="AtLI6246.1"	
/db_xref="GI:16226734"	
/translation="MEAVIQTRGLSLPTKPLGVRSOLOPSHGLKORLFAAKPRNLHG LSLFNGHKKFQFEPTLHGISISHKRSFEKSKAEAAAGCGAGGEDSAVAAS PKIIGVEVATIKRIIPGLMPECLIFNTLTITKTVLIVATAGSSAEIIPKLTWN LPMATGFMILYTKLSNVLSKALFTYVIYPIETIYRAGFVYVPLSNYIHPELAKVL LITLGRFEMPIALIRIMSCLTLYVAELMGVSVVSLTNGFANQITTYDEAKKFP LFGANVALIFSGRTVYKFSNLRKILNGPVGDAVSLKAMSLIVMGALICLLYMW NRYVPLPTRSKNEKPKMGTMESLKLVSSEPIRDLATLVAVGISINLVEYTKSK LKQFSPNEYSAFMGDFSTCVATFTFVLMLLISOYENKVGWVAKITPTVTLVGLG AFPSLILFGSPAPFAVKILGMPILAAVVVGAATLONFESKASYSLEDPCKEMAVITLD EDKRYGKAIDIVCNPLKSGSGLAIQQRHILSFGSLANSTPLGLITLIVLYTAMLA AASLEGFNSLRSSELEKEMERASSVKIPVSDSGNGSGESPSSEKSAPTNL "	
1931. 2110	
3'UTR	
BASE COUNT	
ORIGIN	
Query Match	

Best Local Similarity 54.5%; Pred. No. 1,4e-73;	
Matches 769; Conservative 0; Mismatches 616; Indels 27; Gaps 2;	
Qy	121 TAAGAAAGTTCGCGCAATGCTTCTAATGTTCTCTGATTAACATTAACATGAGGTG 180
Db	393 TGAATAAGATTATCCCTTGGATTGATGATGTTCTTGTATTTCTTAAATACCAATTC 452
Qy	181 TACCGATACAAAGACACTTATATGTGGAGCTCCGTGTCTGTGACAGAGCAATAC 240
Db	453 TGAGGATACAAAGAGATGTCTGTGGTGAACGGGAAGAAAGATTCGAGATTTATC 512
Qy	241 CTTTCATCAAGTTTGGCTGTTGTCCTCCCTGCTGCTATATCTTATCTTATATCA 300
Db	513 CTTTCTGAAGACTTGGGTGATCTTCTATGCGCAATGGGTGTTATCTCTCTACACT 572
Qy	301 AGCTAAGTAATATTTAAGTAGACAGCCCTATTTATGACGGTGAAGCCGCTTTTAA 360
Db	573 AACTCTCAATGTTCTCTCCAGAGAGCTCTGTTTACACTGTTATGTCCCTTCTCA 632
Qy	361 TTTTCTTGGCCCTGTCCGCACTGTAATTTATCCGCTACGAGATGTTTACATCTAC 420
Db	633 TCTACTTTGGGACCTTGTGGTTGTCATGTACCTCTCAGCACTATATTCACCCGAG 692
Qy	421 AATTGCTGACCGTTTACAGCCATCTCTACCTCCAGATTCGTAGACTGTGCGCAT 480
Db	693 CTCGCGAGATTAAGCTCTTACAAACCTCGGCCCAAGATTCAGGTCTTATGCAAT 752
Qy	481 TAAGAACTGACATTTTGCATATTTATGACTTGTGGAACATTAAGGAAAGCTCAT 540
Db	753 TCGGATTTGGAGTTCTCTGTTGTTATGTATGTATGCTGAGCTTGGGGATGTTG 812
Qy	541 TATCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATCCAGAAAGCAAGCTT 600
Db	813 TCTCAGTTCTCTTGTGGGCTTGTCTAATCAGATCACAAGTGTGATGAGCCAAAT 872
Qy	601 TCTACGCTCTTTCGTATGGAGCTAATATTTCTTACTAGCTTCTGTGCTGCAAT 660
Db	873 TCTATCTTGTCTGAGCTTGAGCCAAATGTGACGATTTTCTCAGAAAGCAAGCT 932
Qy	661 TTTGGGCTCAAAAGTATGAGAGCTCCCTTCTGAAGGTGATAGATCTTGGGAAAT 720
Db	933 AATACCTTCTAATCTAGAGAAAGATCTTGTCTCTGAGTTGAGCGCTGGGCACTT 992
Qy	721 TACGCTTTTGTGAGCTATGACTATTTGATCTGAGCTTGTCTTATGCGCAGTTACT 780
Db	993 TGAAGCCATGATGAGCATTTGGTGGAAATGGAGCTCGCCATTTGCTCTCATTTG 1052
Qy	781 GATCATATAGAACGTATTTGACGATCTCTCTATATCCAGAAAGAAATGCAAAAG 840
Db	1053 GGGTCAATA-----GATATGTTCTCTTCCAAACCCGTAGCA 1088
Qy	841 GGAATAAGGTGCTAAGCTAATAATGAATGAAGATAGCTTCTATCTTGATGAT 900
Db	1089 AGAACAAAGAGAGAAACCGAAGATGGGAACGATGGAAGCTTGAAGTTCTGTGAT 1148
Qy	901 CTCCTTATTTCTTATTAATCTCTCTTGTATTTGCTTATGCTTATTTGATTAAC 960
Db	1149 CACCATCATTTAGAGATCTTCTTATTTAGTGGGATACGGATATAGTAAATCTTG 1208
Qy	961 TCGAAGTACTTGGAAAGTACGCTGAACCTCAATATCTTAATATATAGCTATATG 1020
Db	1209 TGGAAAGTACATGGAAATTCAAAGCTTAAACCTCAGTCCCTACCCGAAAGATAC 1268
Qy	1021 AGTTTCATGGGAAGCTTCTCTCTGAGCTGGCGAGTATGATCTATATCTATTTG 1080
Db	1269 CATTTATGAGAGACTTCTCAACCTGCACGGGTGTTGCAACATTCACATATGAT 1325
Qy	1081 TTGCTGTACGTCATTCGTAATTTGATGATGTTAACTGAGCCCTGACTCTGTCA 1140
Db	1326 TCAAGCAATACGATTCATTAATGAATGATGTTGGGAGATGCTGCAAGATCACCCAA 1385
Qy	1141 TCGTTCTCTAAGCAGTATGCTTTTCTGCTCTTGTATCTTTAGAACCAAGCTTCT 1200

Db 1386 TTCTGCTATTCAGTGTGTCGCTTCTTCTCTAATATTTGTTGGCGCCCATTCGAC 1445
 QY 1201 GGCTGGTGCATTCGTTGGTACACCTCTCATGCTGTGCTGGTGGGAGCTATAC 1260
 Db 1446 CACTGTGTGCACAGCTGGTATGACACCGCTACTGACGCTGTATGCTGGTGGCTTC 1505
 QY 1261 AGAATATTCCTTCGAAATTCACAAATACGCTCTCTTTGGCTCACTAAGAAATGGCT 1320
 Db 1506 AGAATATTCCTTCGAAATTCAGAGTGCACAGCTGTGTCACCTTCGAAAGAAATGGCT 1565
 QY 1321 ATATCCCTCTTCAGCAAGACGAAAGTCAAGGTAGAGCTCTATTCATGATGTTGCCG 1380
 Db 1566 ATATCCCTCTTCAGCAAGACGAAAGTCAAGGTAGAGCTCTATTCATGATGTTGCCG 1625
 QY 1381 CCCGCTTCGAAATTCAGAGGAGCTTTATACCAAGAGCTTGCCTGATTCCTGGA 1440
 Db 1626 ACCCATTCAGAAATTCAGAGGAGCTTTATACCAAGAGCTTGCCTGATTCCTGGA 1685
 QY 1441 GTATTCAGAGCTATTCAGAGCTTATTCAGAGCTTCTTTTCATTCATTCATTCGCT 1500
 Db 1686 CACTAGCGAATTCACAGCGCGTATTCAGAGATGATCTGTGTTATTCATTCATTCGCTGCT 1745
 QY 1501 TGGTTTCGCACTAAGTTAAACAACATATTC 1532
 Db 1746 TAGCTGACAGCTAAGTCGCTGAGGAGACAGTTC 1777

RESULT 12

LOCUS ATANTMR 2181 bp mRNA linear PLN 26-NOV-1999
 DEFINITION A. thaliana mRNA for adenine nucleotide translocase.
 VERSION 249227.2 GI:6469339
 KEYWORDS adenine nucleotide translocase.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana;
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2181)
 REFERENCE
 AUTHORS Neuhaus, H.E.
 TITLE Molecular characterization of an Arabidopsis thaliana cDNA encoding
 a novel putative adenylate translocator of higher plants
 FEBS Lett. 374 (3), 351-355 (1995)
 MEDLINE 96069943
 PUBMED 7589569

REMARK (sites)
 REFERENCE 2 (bases 1 to 2181)
 AUTHORS Kampfenkel, K.K.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-1995) Kampfenkel K.K., Universiteit Gent,
 Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium,
 B-9000 Gent
 revised by [3]
 REFERENCE 3 (bases 1 to 2181)
 AUTHORS Kampfenkel, K.K.
 TITLE Direct Submission
 JOURNAL Submitted (26-NOV-1999) Kampfenkel K.K., Universiteit Gent,
 Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium,
 B-9000 Gent
 On Nov 27, 1999 this sequence version replaced gi:1051108.
 COMMENT
 FEATURES
 source
 location/Qualifiers
 1..2181
 /organism="Arabidopsis thaliana"
 /strain="var. Columbia"
 /db_xref="taxon:3702"
 /clone="C23"
 /tissue_type="whole seedling"
 /clone_lib="lambdaYES"
 /dev_stage="seedling"
 110..1981
 /codon_start=1

/product="adenine nucleotide translocase"
 /protein_id="CA89201.2"
 /db_xref="GI:6469340"
 /db_xref="SPTREMBL:Q39002"
 /translation="MEAVTQIRGLSLPTPKDGYRSQLQPSHGLKORLFAKPRNLHG
 CLVPLGTGRNRPPLSPQCMGRFPTRKRSYVRRRRGRSCLRRSDAAVVASR
 KIFGEVATLKKIIPGLMFPCILEFNTIILDDTDVLTVCAGSAEIPILKTMVNL
 PMAFGLMLYTKLSNLVSKLAFYIVPPIIIRGSGFPWYPLSNTHPEALDKLL
 TTVGRPRGRPLAIRIRMSFCLFYMAELMGSVSVLPWGRANDITVDEAKRTPLP
 GLGANVALIFSGRTVKYFSLNRKNGLPQVDMAVSLAAMSIVVGMGLALICLVWVN
 RYVLPPTSRNKKRKPMTGMSLKFVLSVPIIDLATLVVAIGISINLVEVTKSKL
 KAQPPNMEYAPAFMGDSCTCGVATFTMMLSQYFNKGVAGAKITPVLTLTGYA
 EFSLIFEGPAPLAVKGMTPLAAVYVGLVQNFSAKYSLEDPCKEAVYLPDE
 DTKYKRAIDVNCNPLKSGGALILQDMPLISFGINSFPLGIMILLIVYATLAA
 KSLRGFMSLSEERLEKEMRBRASVKIPVYQDSGNGSLGESPSSEKSAFTNL"
 BASE COUNT 549 a 476 c 508 g 648 t
 ORIGIN

Query Match 21.5%; Score 352.4; DB 8; Length 2181;
 Best Local Similarity 54.5%; Pred. 1.3e-73;
 Matches 769; Conservative 0; Mismatches 616; Indels 27; Gaps 2;

QY 121 TAAAGAAAGTCTGCCAATGTTCTCAATGTTCTCTGATTTACATTTAATACGCTGT 180
 Db 444 TGAAGAAAGATATTCCTTAGATGATGATGTTCTTTGTATCTTTCAATTCACAAATTC 503
 QY 181 TACGGATACAAAGACACATCTTATTTGAGAGCTCTGCTGCGACAGGCAATAC 240
 Db 504 TGAGGATACAAAGAGATGCTTGTGGTGACGCGCAAGAAAGTCTGCTGAGATTAAC 563
 QY 241 CTTTCATCAGATTTGGCTGTTGTCCTCCCTGCTATTAATCTTATTCATTTATTCGCA 300
 Db 564 CTTTCGTAAGACTTGGGAGATCTTCTCATATGCGCATTTGGGTTTATCTCTCTACACTA 623
 QY 301 AGCTAAGTAATATTTTAAGTAAGAGCGCTTATTTATGACAGTGGAGAGCGCTTTTAA 360
 Db 624 AACTCTCCACATGTCCTCCACAGAGAGCGCTGTTTATACCTGTATTTGTCCTTCATCA 683
 QY 361 TTTTCTTGCCCTGTTCCGACGTAAATTTATCCGCTACCGCATGTTTACATCTACAG 420
 Db 684 TCTACTTTGGGGCTTTGGTTTGTGATGATGATACCTCTCGCAACATATATCACCGGAAG 743
 QY 421 AATTGCTGACCGCTTTTACAGCGCATCTACCTCCAGAGATGCTAGAGCTGCTGCATCT 480
 Db 744 CTTCTGCAGATTAAGCTCTTACAAACCTCGGCCCAAGATTCATGGGTCCTATTTGCAATAT 803
 QY 481 TAAGAAATGACATTTGCTGATTTTATGATCTTCTGTAACATATGAGGAGCGTCATGC 540
 Db 804 TCGGATTTGGAGTTCTGTTGTTTATGATGCTGAGCTTTGGGCTAGTGTGTGG 863
 QY 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAATTCACAGCAAGCAACCGTT 600
 Db 864 TCTCAGTTCTCTCTGGGGCTTTGCTAATCAGATACCACTGTGATGAAGCAAGAAAT 923
 QY 601 TCTACGCTCTTTTCTGATACGAGCTAATATTTCTTACTACCTCTGCTGCAATTTG 660
 Db 924 TCTATCCTTTGTTGCGAGCTTGAAGCAATGTGCACATGATTTTTCAGAGAAACCGTCA 983
 QY 661 TTTGGCTTCAAGATTTGACAGCTTCCGTTTCTGAAGGTGTAGATCCCTTGGGAATTTCT 720
 Db 984 AATACCTTCTCACTTCAAGAAAGAAATCTTGCTCGCATGATACGCGCGGCACTTTCGT 1043
 QY 721 TACGCTTTTGAATGCTATGATATGATATGATGAGCTGCTTATGAGCAATTTACTGCT 780
 Db 1044 TGAAGCCATGATGACATTTGTGTGGAGATGAGAGCTGCCCATTTGTCTCTCTATTTGGT 1103
 QY 781 GGATCAATTAAGAACGATTTGACCGATCTCGCTTATATATCCAGAAAGATGCAAAAG 840
 Db 1104 GGGTCAATA-----GATATGTTCTCTTCCAAACCGCTAGCA 1139
 QY 841 GGAAGAAAGCTCTAACCCTAAATGATATTAAGATAGCTTCTCTATCTTGATAGAT 900
 Db 1140 AGAACAGAAAGAGAAACCGAAGATGAGAACGATGAGAAAGCTTGAAGTCTTGTGATCAT 1199

OY	901	CCCCATATCTCTTTATTAACCCCTCTGGTATATCCCATGGGATTTGCATTAACCTTAA	960
Db	1200	CACCATACATTAGAGATCTTGGCTACTTTAGTGGTGGCATACGGTATTAGATAACTTGG	1259
OY	961	TGCAAGTACTTGGAAAGCTGCAAGCTGCATATATCTTAATATGAATGACTATATAGT	1020
Db	1260	TGGAGATCACAATGAAATCAAAAGCTTAAAGCTCAGTTCCTAGGCCGATGAOCTACAG	1319
OY	1021	ACTTCATGGGGAACCTCTCTCTTGACATGGGGTASTATCCGACTTATCATGCTATTGG	1080
Db	1320	CATTATGGGAGACTTCTCAACCTGACGGGTGTGGCAACATTCACAAAGATGAC---TTC	1376
OY	1081	TTGGTGGTACGCTATTCGTAATTTGGAATGGTTAACTGAGCCCTAGTACACCTCTGTCA	1140
Db	1377	TCAGCCAAATACGTTATTCATTAAGTATGTTGGGGAGTACTGCAAAAGATCACCCCACTG	1436
OY	1141	TGGTTCCTCAACAGGTATGTTTCTCTGCTCTGTGTAATCTTTAGAAACAAGCTTCTG	1200
Db	1437	TTCTGCTATTTGACTGGTGTGGCTCTTCTCTTAATTTGTTTGGCGCCCATTCGCAC	1496
OY	1201	GGCTGGTGGCTATGTTCGGTACACTCCCTCTCATGCTACTCTGGTGGTGTGGAGCTATAC	1260
Db	1497	CACGTGTGGCCAGCTGGTATGACACCGCTACTTCACACTGTGATATGTCGGTGGCCCTTC	1556
OY	1261	AGAATATTTCTTTCGAAATCCACAAAATCAGCTCTCTTCTACTCAACTAAGAAATGGCT	1320
Db	1557	AGAAATATCTTCAGCAAGAGTGCACAGTACAGCTTGTTCGACCCCTTGCAAAAGAAATGGCT	1616
OY	1321	ATATCCCTCTTGACCAAGACGAAAGATCAAAAGTAAAGGTAGGCTCTATTTGATGTAGTCCG	1380
Db	1617	ATATATCCCATTTGGATATAGGACACCAAGAGTTAAAGGCAAAACCTCGATGTACGTGGTGTGCA	1676
OY	1381	CCCCCTTGGGAAATCAGAGGAGGCTTTATATCCCAACAAGTGTTCGCTGATATCTGGGGA	1440
Db	1677	ACCCATTAGGGAATCAGGGGAGCTTTATATCAGCAGCTTCATGATCTTATCTTTTGGAT	1736
OY	1441	GTATTTGAGACTATGACCCCTTATCTTGCAGTATTCCTTTTTCATCATTAATGCTATTTGGT	1500
Db	1737	CACATGCAATTTCAACGCCGCTATCTAGGAATGATCTTGTGGTTATTTGTCACATCGGTGT	1796
OY	1501	TGGTTTCTGCAACTAAGTTAAACAACATATTC	1532
Db	1797	TAGCTGACGCTTAAGTGGCTGGAGGAGCAAGTTC	1828

LOCUS	AY045903	2151 bp	mRNA	linear	PLN 24-APR-2002
DEFINITION	Arabidopsis thaliana putative adenine nucleotide translocase (At1lg80300) mRNA, complete cds.				
ACCESSION	AY045903				
VERSION	AY045903.1	GI:15028092			
KEYWORDS	FLI_CDNA.				
SOURCE	Arabidopsis thaliana.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 2151)				
AUTHORS	Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,T., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinzaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.				
TITLE	Arabidopsis Full Length cDNA Clones				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2151)				
AUTHORS	Yamada,K., Liu,S.X., Pham,P.K., Banh,T., Dale,J.M., Goldsmith,A.D., Jiang,P.X., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G., Yu,S., Quach,B.L., Carninci,P.,				

TITLE	JOURNAL	COMMENT
<p>Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamlya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Kosena,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,H., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.</p> <p>Direct Submission</p> <p>Submitted (10-Jul-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA</p>	<p>RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RafL cDNAs (RafL cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamlya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.</p>	<p>The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RafL cDNAs: Yamada,K., Liu,S.X., Pham,P.K., Banh,J., Dale,J.M., Goldsmith,A.D., Jiang,P.X., Lee,J.M., Onders,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G., Yu,S., Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Kosena,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.</p>
FEATURES	<p>Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC) contributed equally to this work as PIs.</p>	
source	<p>Location/Qualifiers</p> <p>1. .2151</p> <p>/organism="Arabidopsis thaliana"</p> <p>/db_xref="taxon:3702"</p> <p>/chromosome="1"</p> <p>/clone="RAF107-08-j19 (R10784)"</p> <p>/note="this clone is in a modified pBluescript vector (pLC-1) as a BamHI/XhoI insert.</p> <p>ecotype: Columbia</p> <p>1. .2151</p> <p>/gene="Atlg80300"</p> <p>1. .27</p> <p>/gene="Atlg80300"</p> <p>28. .1902</p> <p>/gene="Atlg80300"</p> <p>/codon_start=1</p> <p>/evidence=experimental</p> <p>/product="putative adenine nucleotide translocase"</p> <p>/protein_id="AAK6577.1"</p> <p>/db_xref="GI:15028093"</p> <p>/translation="MEAVTQIRGLSPTRIGVRSQLOPSHGILKORLFAKPRNLHGILSLFNGHKKRQEPETPLHGISISHKRSRFFICKAAADAAAGDVGESDAAVNSAPKIGVEAATLKLIIPGLMPECFITPTLITRDQDVLVAVYAKSSMIIIPKLTWNLPKMGIEMLYTKLISNVLSKALPFLYIVPITIEYFGAFGVMTYSHIHPELADNLTITGPRMGPIALIRISRECLFYVAMELMSVSVSLPFGAQITTVDEAKKFTPLFGIDGNALLISGRTVAFYSNLRKNIPLGVDAVNSLKAMSVVGGWGLAICILYVWVNRVPLDPTNRKNKKEPKMGIMESLKFVSSPYRLDATTIVAGISINIVETWKSRLKAPSPNEXSAFMGDSCTGVATFTMMILSQYVNNKYGVAAKITPTVLLITGVAAFLSIFEGGEPAFLVAKIGMTPLLAIVVYGAIONIFSRAKYSLEPCKEMAYIPLDIEDTVKRAADIVYCNPLKSGGALLIQPFMLISGSLANSPTPIGLHILYIVANMLAAAKSLDEGFNSLRSEELKEKEMERSVKKIPVSDSESGNSLGESSPSSPEKSAPTNL"</p>	
5'UTR		
CDS		
gene		
misc_difference		
3'UTR		
misc_difference		
BASE COUNT	553 a 451 c 510 g 637 t	
ORIGIN		
Query Match	21.4%; Score 350.8; DB 8; Length 2151;	

Best Local Similarity 54.4%; Pred. No. 3,36-73;
Matches 768; Conservative 0; Mismatches 617; Indels 27; Gaps 2;

121 TAAAGAAAGTTCTGCCAATGTTCTCTAATGTTCTGATTAACATTTAATACATGCTGT 180
 Db TGAAGAAAGTTCTGCCAATGTTCTCTAATGTTCTGATTAACATTTAATACATGCTGT 180
 365 TGAAGAAAGTTCTGCCAATGTTCTCTAATGTTCTGATTAACATTTAATACATGCTGT 424
 181 TAGCGATACAAAGACGCTGTTATGTTGAGGCTCCGTTGCTGCTGACGACGCAATAC 240
 Db TGAAGAAAGTTCTGCCAATGTTCTCTAATGTTCTGATTAACATTTAATACATGCTGT 424
 425 TGAGGATACAAAGACGCTGTTATGTTGAGGCTCCGTTGCTGCTGACGACGCAATAC 484
 241 CTTTCATGACATTTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 Db CTTTCATGACATTTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 485 CTTTCATGACATTTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 544
 301 AGCTAAGTAATATTTTAAAGACAGGCTTATTTATGACAGTGGGAAGCCCTTTTAA 360
 Db AACTCTCAATGTTCTGCTCAAGAGGCTGTTTACAGTGTATGTTGCTGCTGCTGCTGCT 604
 545 AACTCTCAATGTTCTGCTCAAGAGGCTGTTTACAGTGTATGTTGCTGCTGCTGCTGCT 604
 361 TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
 Db TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
 605 TCTACTGTTGGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 664
 421 AATTGCTGACGCTTTAGAGGCAATCTACCTCCAGATTGCTAGAGCTGTTGCTGCTGCT 480
 Db CTTTCATGACATTTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 724
 665 CTTTCATGACATTTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 724
 481 TAGAAGCTGACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 Db TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 725 TGCGGATTTGAGTTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTG 784
 541 TATCTCTATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600
 Db TATCTCTATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600
 785 TCTCAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 844
 601 TCTCAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
 Db TCTCAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
 845 TCTCAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 904
 661 TTTGGGCTTCAAGTTGAGAGCTTCCGTTTCTGAAAGTGTAGATCCTTGGGGAATTTCT 720
 Db TTTGGGCTTCAAGTTGAGAGCTTCCGTTTCTGAAAGTGTAGATCCTTGGGGAATTTCT 720
 905 AATATCTCTCTAATGTTGAGAAAGATCTGTTGCTGAGATTGACGGTGGAGCATTTGCT 964
 721 TACGCTTTTGAATGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 780
 Db TACGCTTTTGAATGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 780
 965 TGAAGCCATGATGAGCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1024
 781 GATCAATATGAAAGCTATTTGACCGATCTGCTCTATATATCCAGAAAGATGCAAAAG 840
 Db TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 1025 GGGTCAATG-----GATATGTTCTCTCTGCAACCGTAGCA 1060
 841 GGAAGAAAGGTTGCTAATGCTAATGATGAAAGATGCTTCTCTATCTTGTATGAT 900
 Db TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 1061 AGAACAAAGAGGAAAGCAAGATGAGGAAAGCAAGATGAGGAAAGTTCTGTTGATCAT 1120
 901 CTCCTTATATTTTATTAATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 960
 Db CTCCTTATATTTTATTAATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 960
 1121 CACCTATATTTGAGATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1180
 961 TCGAAGTACATGGAAGATGCTGGAATGCAATATCTAATGATGATGATGATGATGATG 1020
 Db TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 1181 TGAAGATGATGGAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAG 1240
 1021 AGTTGATGGAAGATGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 Db AGTTGATGGAAGATGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 1241 CATTTATGAGGAGATCTTCAACCTGAGGAGGTTGCAACATTTCAACATGATGCTG 1297
 1081 TTGTTGTTGATGCTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTG 1140
 Db TTGTTGTTGATGCTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTG 1140
 1298 TCAGGCAATACGTTATTTCAATAGTATGTTGGGAGTACTGCAAGAACACCCCAACTG 1357
 1141 TGGTCTCTCTAAGGATGCTGTTTCTGCTGCTGTTTCTGTTTCTGTTTCTGTTTCTG 1200
 TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG

Db 1358 TTCTGCTATGACGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 1417
 Qy 1201 GCGTGTGCTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Db 1418 CACTGTTGCAAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1477
 Qy 1261 AGAATATCTTCTGCAAAATCCCAAAATACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 Db 1478 AGAATATCTTCTGCAAAATCCCAAAATACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1537
 Qy 1321 ATATCCCTCTTGTACCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 1380
 Db 1538 ATATCCCTCTTGTACCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 1597
 Qy 1381 CCGCTTGGGAAATCAGAGAGCTTTAATCCAAAGGTTGCTGCTGCTGCTGCTGCTGCTG 1440
 Db 1598 ACCATTTGAGGAAATCAGAGAGCTTTAATCCAAAGGTTGCTGCTGCTGCTGCTGCTG 1657
 Qy 1441 GATTTGAGGCTATGACCCCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 Db 1658 CACTAGGCAATTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1717
 Qy 1501 TGGTTTCTGCAACTAAGTTAAACAACATATTC 1532
 Db 1718 TAGCTGCAAGTAAAGTCTGCTGAGGAGCAAGTTG 1749

RESULT 14
 LOCUS AY128844
 DEFINITION Arabidopsis thaliana putative adenine nucleotide translocase
 ACCESSION AY128844
 VERSION AY128844.1 GI:22136331
 KEYWORDS FLI-CDNA.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
 1 (bases 1 to 1896)
 Tripf, M., Southwick, A., Karlin-Neumann, G., Nguyen, M., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H., Kawai, J., Kim, C., Lin, J., Liu, S.X., Matsuda, M., Pham, P.K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shin, P., Yamada, K., Shinzaki, K., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (01-JUL-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

TITLE
 JOURNAL
 COMMENT
 e-mail for correspondence: arab@sequence.stanford.edu

The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: "RIKEN Arabidopsis Full-length cDNA") Seki, M., Natsumura, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinzaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAFL cDNAs: Tripf, M., Nguyen, M., Southwick, A., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shin, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Tripf, M. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinzaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

LOCATION/Qualifiers
 1. 1896
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"

```

/chromosome="1"
/clone="024754"
/notes="This clone is in pENTR/SD-dTopo. This is a cloned
PCR product using RIKEN clone RAFL09-64-M24 (AY081350) as
a template
ecotype: Columbia"
gene
1.1896
/genes="Atlg15500"
/notes="T16N1.1"
1.1857
/genes="Atlg15500"
/codon_start=1
/product="putative adenine nucleotide translocase"
/protein_id="AA091244.1"
/db_xref="GI:22136332"
/translation="MEGLIOTRGILSLPAKPIGVRLLOPSHGLKORLFTNLPALSL
SSNGHKRFAOFIOICISYSHKERSNGFICKAEAAAGAGNFPDEDDTAAAMVSPKI
FVEVVTLLKIVPLGLMFCILNRTITLRTKVLVYTAAGSSAEIIPFLKTVNLPIM
AIGFMLYTRKLSNVLSKALFVIVPEIYFAGFVMPPLSNLHPKALADKLAT
LGRFNGPLAIRIWSFCLFVYAAELMGSVVSVLEMGANOITLYDEARKFPLFL
GANVALIFSGRTVYKESNMKNLPGVDGVAISLAKMSIVMGALICFLYMWVNY
VPLPTSRKKKRVKPMQGTMSLKEFLVSPYIRDLATLVAYGKISINLEVTMSKLA
QEPSPNEYSAFMGDEFSTGCIATFTMMLLSQYKRYKGMVAARIPTVILLGVAF
SILRSGPAPLPAKLGMPPLLAUVGAQNIFFSKAKSLDEPKEMAYPLDEPT
KVGKAIDVNCNPLKSGGALIQOEMILITGSLANSTPYLGLILITVYAMLAAS
LEQFNTLMSEELEREMERASSVKIPVVSQEDAPSGEITTSQISEKSTPTGI"
BASE COUNT 474 a 415 c 456 g 551 t
ORIGIN
Query Match 20.6% Score 336.6; DB 8; Length 1696;
Best Local Similarity 54.0%; Pred. No. 8.1e-70;
Matches 752; Conservative 0; Mismatches 614; Indels 27; Gaps 2;
QY 120 CTAAGAAAGTTCGCCATGTTCCTATGTTCTGTATTAACATTAACGCTG 179
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 328 CTGAAGAAAGATGTTCTTGGGCTATGTTCTTTCATCCTTTCAATTAACATC 387
QY 180 TTACGCATACAAAGACACTCTTATGTGGAGCTCCTGTTCTGTGCAGAGCAATA 239
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 388 CTTAGGACACAGAGATGTTTGGTGTGACGCTAAGAGATGCTGCGAATTTATA 447
QY 240 CCTTTCATCAAGTTTGGCTGTGTGCCCTGTGCTATTAATCTTATGCTATTATGCA 299
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 448 CCTTTTGAAGACATGGGTAAATCTTCGATGCTATGGGTTATGTTCTATACACC 507
QY 300 AAGCTAAGTATATTTTAAGTAAGCAGGCTTATTTATGAGTGGGAACGCCCTTTTAA 359
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 508 AAATTTCCATGTTCTCTCCAAAGAGCTCTTTTACATGATATTTGTTCTTTGATT 567
QY 360 ATTTTCTTGGCCGTGTCCGACTGTATTTATCCGCTACGCGATGTTTACATCTTACA 419
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 568 GTCTACTTGTGGAGCTTGTGGTGTGATGATCCCTCAGCAATTTGATTCATCTGAA 627
QY 420 GAATTTGCTGACGTTTACAGACATCTACCTCAGAGATGGTAGAGCTGTGGCCATC 479
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 628 GCTCTTGTCTGAATGCTTCTTGCAACACTGGCCCAAGATTCATGGGCTCTCTGCAATC 687
QY 480 TTAGAAACTGACATTTGCTGCTATTTATGCTTGTGCTGCTGCTGCTGCTGCTGCTG 539
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 688 ATGAGATTTGGAGTTCTGTTGTGTCATGTCATGGCTGAGCTTGGGAGTATGTTGTC 747
QY 540 CTATCTCTATGTTCTGGGAGTTTGTCTAATTAACAAATTCACGAAGCAAGCGT 599
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 748 GTTTCAGTTCCTTCTGCGGATTTGGCCAAACAGATTTACACTGTTGAGAGCAAAAAG 807
QY 600 TTCTACGCTCTTTCGCTATGCTGAGCTATTTCTTACAGCTCTGCTGCTGCTGCTGCT 659
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 808 TTCTATCTCTGTTTGGCTTGGGCGCAATGTGTGACTTATCTTCTCGAGGAACCTGTG 867
QY 660 GTTGGGCTTCAAGTTGAGAGCTTCCCTTCTGAGAGTGTAGATCTTGGGGAATTTCT 719
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 868 AAATATTTCTCTAATATGAGAAAGATCTTGCTCTGAGATTGATGCTGGGCTGTTTCA 927

```

```

QY 720 TTACGCTTTTGTATGCGCTATGCTATGATCTGCACTGTTCTTATGCGCACTTACTCG 779
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 928 TTAAAGCTATGAGTATGCTATGCTGGGAGTGTCTCCGCAATCTGTTCTCTACTGCG 987
QY 780 TGCATCAATAGAACGATTTGACCGATCCTCGCTTCTATATCCAGAAAGATCCAAAG 839
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 988 TGGGTGATA- - - - - - - - - - - - - - - - - - - - - - - - - - - - -
QY 840 GGGAAAAAGGTGCTAAACCTAAATGAATATGAAGATAGCTCTCTATCTTATGATA 899
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1024 AAGAAGAAAGATGTAAGAACCCACATGGGAACATGAGAGCTTGAAGTTCTTGTTGCA 1083
QY 900 TCTCTATATTTCTTTATTAATCTCTCTGTTATTCCTATGCTATTTGATTAACCTTA 959
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1084 TCACCATATCATTAAGATCTGTACTTGTGCTGCTCATATGAATATGATATCAACCTT 1143
QY 960 ATGCAAGTACCTTGAAGAAAGTACAGCTGAACCTGAATATCTAATATGAATGATATAGT 1019
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1144 GTTAAGTACATGTAAGTAATCAAAAGCTTAAGCTAGTCCCTACCCGAGAAATACTCA 1203
QY 1020 GAGTTTATGGGAACTTCTCCTCTGAGACTGGCTAGTATCCGCTATTCATCATCTATTT 1079
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1204 GCATTTATGGGCGACTTCTCAACCTGACAGAGTATGGAAC- - - ATTCACAAATGATGCTT 1260
QY 1080 GTTGTGTATGCTATCTATCTGTAATTTGATGTTTAACTGAGAGCCCTAGTACTCTGTC 1139
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1261 CTAAGCCAAATACGTGTTAAGAGATGATGTTGGGAGTAGCTGCAAAAGATCAGACACACC 1320
QY 1140 ATGCTTCTCTTACAGAGTATGCTTCTGCTCTGCTTGTATCTTATGAACCAAGCTTCT 1199
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1321 GTTCTGCTATTTAGACCGGTGTGCTTCTCTCTGATGATGTTGGTGGCCCATTTGCA 1380
QY 1200 GGGCTGTGCTATGTTTGGTATGACATCTCTCATGCTAGCTGTGTTGGAGCTATA 1259
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1381 CCAATGGTGTGCAAGCTGTGATGACACCGCTACTCGACAGAGTATGAGTGTGGCCCTC 1440
QY 1260 CAGAAATTTCTTGGAAATCACAAGATAGCTCTCTTGAATCTCACTAAGAAATAGGCC 1319
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1441 CAGAAATCTTCCAGACAGATGACAGTACAGCTTGTGCAATCTTGCAGAAATAGGCT 1500
QY 1320 TATATCCCTTGTGACCAAGACCAAAAGTCAAGAGTGAAGCTGTATGATGATGTGGCC 1379
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1501 TATATCCATTTGATGAGAGACCAAGSTTAAAGCAAGCTGCAATTTGATGTGCTGC 1560
QY 1380 GCCCGCTTGGGAAATCAGAGAGAGCTTTAATCCAGAGGTTTGTCTGTTATCTGTGA 1439
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1561 AACCATTGGGGAATATCAGCGGTGCTTAAATCCAGAGATTCATGATGCTTAACTGGCC 1620
QY 1440 AGTATTTGAGCTATGACCCCTTATCTTGCAATGTTCTTTCATCATTTGATTTGG 1499
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1621 TCACCTGCGCAATTCACACACTTACCTTGAGATCATCTGCTGAGTATGATGCTGCAATG 1680
QY 1500 TTGGTTCTGCAA 1512
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1681 TTACAGCAGAGCTA 1693

```

RESULT 15
 LOCUS AY081350
 DEFINITION Arabidopsis thaliana putative adenine nucleotide translocase
 (Atlg15500) mRNA, complete cds.
 ACCESSION AY081350
 VERSION AY081350.1 GI:19699006
 KEYWORDS F1.CDNA.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 2090)
 Nguyen,M., Karlin-Neumann,G., Southick,A., Lam,B., Miranda,M.,
 Palm,C.J., Bowser,J., Jones,J., Bohn,J., Carninci,P., Chen,H.,

Db 1697 AACCATTTGGGGAAATGAGGGGGCTCTTAATCCAGCAGTTCATCATCTTACATTCGCG 1756
 QY 1440 AGTATGGAGCATGACCCCTTATCTTCAGTATCTTTTCATCATCTTATTCGTTGG 1499
 Db 1757 TCACTCGCCAATTCACACACTTACCTTGAGTATCTCTGCGTATAGTACCTCATGCG 1816
 QY 1500 TTGGTTTCTGCAA 1512
 Db 1817 TTAGCAGCAGCTA 1829

RESULT 16
 AY084374 2146 bp mRNA linear PLN 21-JUN-2002
 LOCUS Arabidopsis thaliana clone 105967 mRNA, complete sequence.
 DEFINITION AY084374.1 GI:21403084
 VERSION FLI_CDNA.
 KEYWORDS thale cress.
 SOURCE Arabidopsis thaliana
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 TITLE 1 (bases 1 to 2146)
 Haas, B.J., Voliovsky, N., Town, C.D., Troupkhan, M., Alexandrov, N., Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.
 Full-length messenger RNA sequences greatly improve genome annotation
 JOURNAL Genome Biol. (2002) In press
 REFERENCE 2 (bases 1 to 2146)
 AUTHORS Brover, V., Troupkhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.
 TITLE Full-length cDNA from Arabidopsis thaliana
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 2146)
 AUTHORS Brover, V., Troupkhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA
 COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Ler ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.
 FEATURES
 source Location/Qualifiers
 1..2146
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /clone="105967"
 113..1969
 /codon_start=1
 /product="adenine nucleotide translocase, putative"
 /protein_id="AAM60955.1"
 /db_xref="GI:21536623"
 /translation="MEGLIOTRGILSLPAKPIVVRRLQPSHELKORLFTTNPAISLSSWGHRKFOAFQDIPGLGISVSHERSGFLCKAEAAAGGANNVEDDTHAAMAVSPKIFGVETLTKRIPGLKMFECILFNNTILRPTKDLVLTAAAGSSAIEIPFKTWNLMPIAIGMLTYKLSLVNLSKALFYIVPEIYFGAFVYGLNLIHDEKAFKPLDLATLGRPMPLAIRMISFCLFYVMAELMGSVVSVLPMGFPANOTJTDDEAKFPLDGLGANVALIFSGRIVKYSNMKNKNGPGVDGNAVSLKAMSLIVMGALICHLVWVNRV

BASE COUNT 533 a 464 c 499 g 650 t
 ORIGIN
 VLEPFRSKRRKRVQWGTMTESLFLVSSPYIRDLATLVAVGISINLEVTWMSKLA
 QPSPNENYSAEMGDFSTCTGIAFTMMLLSQVYFKRYGMVAKITPTVLLTGAVF
 SLIFGSPFAPALVQKIMPLAAVYVLAQVNFISKSKSLDFPCKEAVYIPLEDT
 KVGKALIDVNCNLIKSGALFIDQEFMLNIFLEGSILANSPTPLGILIGIVAMIAAAS
 LEOFNLMBEELEREMERASSKIPVISOEDAPSETTSLSSEKSTPLGI"
 Query Match 20.6% Score 336.6; DB 8; Length 2146;
 Best Local Similarity 54.0%; Pred. No. 8e-70;
 Matches 752; Conservative 0; Mismatches 614; Indels 27; Gaps 2;
 QY 120 CTAAGAAAGTTCGCCAATGTCCTAATGTTCTGTATTTAATTAATTAATGCGTG 179
 Db 440 CTGAAGAAGATGTTCTTTAGGCTATATGTTTTCATCTCTTCAATTAACAATC 499
 QY 180 TTACGCGATACAAAAGACACCTTATGTGGAGCTCCTGTTCTGTGCGAGGCAATA 239
 Db 500 CTTAGGACACGAAGAGATGTTTGTGTGACGCTAAAGGAAGTCTGTGAATATATA 559
 QY 240 CTTTCATCAAGTTTGTGCTTGTGCCCCGTGCTATATATCTTATGCTTATTTATGCA 299
 Db 560 CCGTTTGAAGACATGGGTGAATCTTCCGATGGCTATGGGTTATGTTCTATACACC 619
 QY 300 AAGCTAAGTATTTTAAAGTAAAGCAGGCTTATTTATGCAAGGGAAGCCCTTTTAA 359
 Db 620 AAACCTTCCATGTTCTCTCCAAAAGCCCTCTTTTACATGTTATGTTCTTTCAAT 679
 QY 360 ATTTCTTTGCCCTGTTCCGACATTAATTCGCGACGATGTTTATCATCTGACA 419
 Db 680 GTCTACTTTGAGGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 739
 QY 420 GAATTTGCTGACGCTTTACAGAGCCATCTTACCTCAGAGTTGTAGAGCTTTGCCATC 479
 Db 740 GCTTTGCTGATTAAGCTTTCTGCAACACTCGGCCAGATTCATGAGGCTCTGCAATC 799
 QY 480 TTAAAGAACTGCAATTTGCTGCAATTTTATGTTACTGCTGCACTATGGAAGGCTATG 539
 Db 800 ATGAGGATTTGAGATTTCTGTTTGTATGTCATGCTGAGCTTTGGGATGTTGTC 859
 QY 540 CTATCTCTAATGTTCTGGGATTTGCTAATGAATTTACAAATTCACGAAGCAACGCT 599
 Db 860 GTTTCAGTTCCTCTCTGAGGATTTGCAACAGATTTACACTGTGGAAGCAACCAAG 919
 QY 600 TTCTACGCTTTTTCGATGAGAGCTAATATTTCTTACTAGCTTGTGCTGCAAT 659
 Db 920 TTTATCTCTGTTTGGAGCTTGGGCAATGTGACATTAATCTTCAGGAAGACGTG 979
 QY 660 GTTGGGCTTAAAGTTGAGAGCTTCCGTTTCTGAAGGTGAGTCTTGGGGAATTTCT 719
 Db 980 AAATATTTCTAATATGAGAAAGAAATCTTGCTCGAGATGAGCTGGGCTGTTTCA 1039
 QY 720 TTACGCTTTTGAAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 779
 Db 1040 TTAAGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1099
 QY 780 TGATCATATAAGACGATTAAGCAGATCTCGCTTATATATTCAGAAAGATGCAAAAG 839
 Db 1100 TGGGTGAATA-----GATATGAGCCCTCCCAACCCGTAAGC 1135
 QY 840 GGGAAAAAGGTGCTAAACCTAAATGAATATGAAGATAGCTTCTATCTTGATAGA 899
 Db 1136 AAGAAAGAAAGGAGGAACCAACAGATGGAACATGAGAGCTGAAGTCTTGAGTCA 1195
 QY 900 TCTCTATATATTTTATTAATACCTCTGTGTTATGCTATGATGATGATGATGATGATG 959
 Db 1196 TCACCATATCATTTAGGATCTTGTCTACTTTGTGTGTTCAATATGATATGATATCAACCT 1255
 QY 960 ATCGAATGATCTGGAAGACGCTGAACCTGCAATATGATATGATATGATGATGATGAT 1019
 Db 1256 GTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1315
 QY 1020 GAGTTCAATGGGAACTTCTCTCTGAGCTGGGCTAGTATTCCTACTTATCATGCTATTT 1079

Db 1316 GCATTATAGGGGAGCTTCACACCTGCACAGGATATGCAAC---ATTCACATGATGCTT 1372
 QY 1080 GTTGGTGGTAAGCTATTCGTAATTTGGATGGTAACAGGCCCTAGCTCTGTGC 1139
 Db 1373 CTAAACCAATAGTGTATTAAGAATATGTTGGGAGTAGCTCAAAAGATCACACCAACC 1432
 QY 1140 ATGGTCTCCCTAACAGATATCTTTCTCGCTCTTGTATCTTTAGAAACCAAGCTTCT 1199
 Db 1433 GTTCTGATTTAGCCGGGTGTTGCTTCTCTCTGATACCTGTTGGTGGCCATTCGCA 1492
 QY 1200 GGGCTGCTGCTATGTTGGTACAACTCCTCATGCTAGCTGTGGTGTCTCGAGCTATA 1259
 Db 1493 CCATTGGTTGCCAGCTGTATGACACCGCTAGCTGACGAGCTGTAGCTGGTGGCCCTC 1552
 QY 1260 CAGATATTTCTTCCAAATCCCAAAATAGCGCTCTCTTTGACTCACTAAGAAATGGCC 1319
 Db 1553 CAGAAATATCTTACAGCAGAGTGCACAAATACAGCTTGTCTGATCTTGCAGAAATGGCT 1612
 QY 1320 TATATCCCTCTGACCAAGCAAAAGTCAAAAGTAAAGGCTGATTTGATGATGCTGCGC 1379
 Db 1613 TATATCCCTCTGATGAGGACACCAAGGTTAAAGCAAGCTGCATTTGATGTGGTCTGC 1672
 QY 1380 GCCCGCTTCGAAATTCAGAGAGAGCTTTATCAACAAGTTCCTGCTTATCTGTGA 1439
 Db 1673 AACCATTTGGGAAATTCAGAGCGGTGCTTATTCAGCAAGCTGATGCTTCAATTTGCGC 1732
 QY 1440 AGTATTTGAGCTATGACCCCTTATCTTGATGATGATCTTTTATATCTGCTATTTGG 1499
 Db 1733 TCACTCGCCAATTCACACACTTACCTTGGAGTATCTGCTGATGATGATGCTGATG 1792
 QY 1500 TTGGTTTCTGCAA 1512
 Db 1793 TTAGCAGCAGCTTA 1805

RESULT 17
 AX098422 1770 bp mRNA linear PAT 02-APR-2001
 LOCUS AX098422
 DEFINITION Sequence 1 from Patent WO0120009.
 ACCESSION AX098422
 VERSION AX098422.1 GI:13537714
 KEYWORDS
 SOURCE
 ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 1770)
 AUTHORS Reindel, A., Geisenberger, P. L., Neuhaus, H. E., Greve-Kampfenkel, K.,
 TITLE Moehlmann, T. and Tjaden, J.
 JOURNAL Plants having altered amino acid contents and method for the
 PATENT: WO 0120009-A 1 22-MAR-2001;
 BASF AKTIENGESELLSCHAFT (DE)
 FEATURES
 SOURCE Location/Qualifiers
 1..1770
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"

BASE COUNT 441 a 383 c 431 g 515 t
 ORIGIN

Query Match 20.2%; Score 330; DB 6; Length 1770;
 Best Local Similarity 54.5%; Pred. No. 3,1e-68;
 Matches 769; Conservative 0; Mismatches 610; Indels 33; Gaps 4;

QY 121 TAAATAAAGTCTGCCAATGCTCTAATCTTCTGATTAATTAATTAACGCTGT 180
 Db 335 TCAAAAAGATTAATCCCTTAGATGATGTCTCTTTGTATCTTTCAATTCACACATTC 394
 QY 181 TACGCGATACAAAGACACTTATTTGAGAGCTCGGTTCTGCTGAGAGCAATAC 240
 Db 395 TGAGGATATCAAAAGATGTCTGTGTGAGCGCAAGAGATTTCTGCTGAGATTATAC 454

QY 241 CTTTCATCAAGTTTGGCTTGTGCCCCGTGCTATTATCTTANCGTTATTATGCAA 300
 Db 455 CTTTCTTGAGACTTGGGTAATCTTCTATGCGCATGGTTATAGCTCTCTACACTA 514
 QY 301 AGCTAAGTAATATTTTAAAGTACAGGCTTATTTTATGACATGGGAAAGCCCTTTTAA 360
 Db 515 AACTCTCCAAATGCTCTCCCAAGAAAGGCTGTTTATACAGCTTATTTGCTTTCACCA 574
 QY 361 TTTTCTGCCCCCTGTTCCCGACTGAATTTATCCCTACGCGATGTTTACATCTACAG 420
 Db 575 TCTACTTTGGGGCTTTGGTTTGTCAATGATCCCTCAGCAAAATATATTCACCCGGA 634
 QY 421 AATTGTCAGCCGTTTACAGGCAATCCTACCTCCAGAGATGCTAGACTCGTTGCATCT 480
 Db 635 CTCTGCGATATAGCTCTCTTACACCCCTGCGCCCAAGATATGAGGCTCTATTTGAAT 694
 QY 481 TAAGAACTGCAATTTGCTGCAATTTATATGCTGCTGAACTATGAGGGAAGCCATGAC 540
 Db 695 TCGGATTTGGAGTTTCTGTTTATTTATGTTATGATGCTGAGCTTTGGGTAGTGTGG 754
 QY 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATCCACGAAAGCCTT 600
 Db 755 TCTCAGTTCTCTTCTGGGCTTGTCTTAATCAGATCAACAACCTGTGATGAGCCAAAG 814
 QY 601 TCTAGCCTCTTGGGATCGGATCGAATATTTCTTTACTACTCTCTGCTGCTGCAATG 660
 Db 815 TCTATCTTGTGCGCATGAGCAATGTGCACTGATTTTCTGAGAAACCGCTGA 874
 QY 661 TTTGGGCTTCAAAAGTTGAGAGCTTCCGTTTCTGAAGGTGTAGATCCTTGGGAAATTTCT 720
 Db 875 AATACTTCTCTAATCTTGAAGAAATCTTGGCTCGAGATTA-----CGGCAATTTGCT 929
 QY 721 TACGTTCTTGAATGCGTATGACTATTTGATCTGACTGTGTTCTTATGCGCACTTACTGT 780
 Db 930 TGAAGCCATGATGAGCAATTTGTGGAGATGAGACTCG--CATTTGCTCTCTATTTGCT 987
 QY 781 GGATCAATTAAGACGATTTAGCCGATCCGCTCTATATATCCAGAAATGCAAAAGG 840
 Db 988 GGGTGAATAGA-----TATGTTCTCTTCCACCCCTGACA 1024
 QY 841 GGAATAAAGGCTTAACCTAAATATGAATGAAGATAGCTTCTATCTTGTATGAT 900
 Db 1025 AGAACAGAAGAGAAACCGAAGATGAGAGCATGGAAGCTTGAAGTTCTTGATATCAT 1084
 QY 901 CTCCTTAATTTCTTTTATTAATCTCTCTGTTATTTGCTATGATTTGCAATTAACCTAA 960
 Db 1085 CACCATATACATTAGAGATCTTCTACTTATAGTGTGCGATACGCTATTTAGTATCAATCTTG 1144
 QY 961 TCGAAGTGACTTGGAAATGCAAGTCAAGTCAATATCCATATATGAATGATATAGT 1020
 Db 1145 TCGAAGTCAATGAGAAATCAAGCTTAAGCTTCACTGCTCCATGCGCGGAATGAGTCACTG 1204
 QY 1021 AGTTATGAGGGAATCTTCTCTGAGAGTGGATGATCCCTACTTATATCATGATATTG 1080
 Db 1205 CATTTATGAGGACATTTCTTAACCTGACGGGTGTGCAACATTCACAAATGATGCT---TTC 1261
 QY 1081 TTGTTGTATACGCTATCTGTAATTTGATGTTAATCTGAGAGCCCTAGTCACTGCTGCA 1140
 Db 1262 TCAGCAATACGATTAATCAATAGTATGTTGGGAGTATGCTCAAAAGATCACCCCAAGTG 1321
 QY 1141 TGGTTCTCTAAGATATGCTTTTCTGCTGCTCTGTATATCTTTAGAAACCAAGCTTGTG 1200
 Db 1322 TTCTGCTATTTAGCTGTGTGGCTTCTCTCTCTATATATTTGTTGGCCCATTTGCGAC 1381
 QY 1201 GCGCTGCTGCTATGTTGCTGCAACTCTCTATCTAGCTGTGTTGGAGGCTATATAC 1260
 Db 1382 CACTTGTGCGCAAGCTGTGATATGACACCGCTACTTGCTGATGTATGCGTGTGCTG 1441
 QY 1261 AGAATATTTCTTGAATTCACAAATAGCCTCTTTGACTCACTAAGAAATGGCTT 1320
 Db 1442 AGAATATCTTACGAAAGATGCGCAAGTACAGCTTGTTCAGCCCTTGCAAAAGAAATGGCT 1501

QY	1321	ATATCCCTCTTGACCAAGAGCAAAAAAGCTCAAGAGTAAGCTGCTATATGATAGTATGCGG	1380
DB	1502	ATATCCCATTTGATGAGGACACCACAGGTTAAAGGCAAAAGCTCGATTTGACGGCTGTGA	1561
QY	1381	CCCCCTTGGGAAATACAGAGGAGGAGCTTTATATCCAACAAAGCTTGGCTGTTATCTGTGAA	1440
DB	1562	ACCCATTTAGGAAATATCAGGGGGAGCGTTTAAATACAGAGTTCATGATCTTATCTTTGGAT	1621
QY	1441	GTATTGAGACTATGACCCCTTATCTTGACGATGATCTCTTTTCATCATATTCATTTGGT	1500
DB	1622	CACGTAGCGAATTCACACGGCGGTATCTAGGAATGATCTTGTGGTTATTTCTACCTGCGTGT	1681
QY	1501	TGGTTCTGCACTAAGTTAAACAAACTATTC	1532
DB	1682	TAGCTGCACTAAGTCTGCTGAGGAGCACTTC	1713
RESULT	18		
AX100499			
LOCUS	AX100499	1823 bp	DNA
DEFINITION	Sequence 1 from Patent WO0121803.		linear
ACCESSION	AX100499		
VERSION	AX100499.1		
KEYWORDS	GI:13619512		
SOURCE			
ORGANISM	Chlamydia pneumoniae.		
REFERENCE	Chlamydia pneumoniae.		
AUTHORS	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
TITLE	1 (bases 1 to 1823)		
	Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.		
	I chlamydia /1 antigens and corresponding dna fragments and uses		
	thereof		
JOURNAL	Patent: WO 0121803-A 1 29-MAR-2001;		
FEATURES	Aventis Pasteur Limited (CA)		
source	Location/Qualifiers		
	1..1823		
	/organism="Chlamydia pneumoniae"		
	/db_xref="taxon:83558"		
	101..1723		
	/note="unamed protein product"		
	/codon_start=1		
	/transl_table=1		
	/protein_id="CAC36447.1"		
	/db_xref="GI:13619513"		
	/translation="MQSSEVKPESRLRAYLCPIKSESKFVPLEFLAFVGNVCLL		
	/KMDTLVIGSDGAEYIPFLKVGIVPGAVITVMYGMGSHRPRDTPVCFMAE		
	LGFPEFLAVIIPVGDGHLNLSADKLOELPQRLRGFIWVRWYSIYVMSLELMS		
	DSVSMLEFMCGLITITTEGREFALINTLSSICAGISVWVGQTFVAEFACG		
	DSHVSVMNLMTLGTGLIMVIRLHHTIDTSTSPSRRLVAREGAATANKKKK		
	KPKARKNLPHLIOSRYLGLAIVISYNVITLPEVMEKDOYSOITSYSHVENGYW		
	SRITTLGLVSVYLAALVLTGCTIRKGTGVALVPLVMLVSGILFEFTTFAARDIS		
	IFPGVLGMPALAAWTGMONVLSRGKTFEPQTEEMATIPLSPEKDNKGAIDG		
	VVSHRIGSGSLIYGLLVLFSSVAASLNVIALVLLIMVIAVIVATIGKEYSRRAA		
	DAVNTLKPKEPSSSIYREAOESYEOEMAVL"		
BASE COUNT	450 a 359 c 396 g 618 t		
ORIGIN			
Query Match	18.4%; Score 301.4; DB 6; Length 1823;		
Best Local Similarity	50.3%; Pred. No. 2.1e-61;		
Matches 814; Conservative 0; Mismatches 776; Indels 27; Gaps 2;			
QY	2	AAATRAAAACTATTCAGATAGAAATATAAAGTATTTCAGAGGTAATATGACAAAAAC	61
DB	52	AACATAAAGTATAAGCTTTATACGCCACATAAAGTGTGGAAAGGATTATGACACTCATC	111
QY	62	CGAAGAAAAACCTTTGGAAATGTCCGCTCTTCTGTGGCGGATACATPACTCAGAGCT	121
DB	112	AGAAGTGAACCCCTTTTCAGAGCTCGCGGCATATCTTTGCTCTATTATTAATCAGAATT	171
QY	122	AAAGAAAGTTCTGCCAATATGTTCTCAATTTCTTCTGATATTTACATTTAATACGCTGT	181
DB	172	TTCTAAGTTGTTCACATTTTCTACTAGCGTTTTTTCGTTGGCTTTTACTACTAGCTGCT	231
QY	182	ACGGGATACAAAGCACTCTTATTTGTGGAGACTCTGCTTCTGGTGCAGAGGCAATAC	241

Db	232	GAAGAAACATGAAGATFACCTGCTGCTCATCTGTGGCTTACAGATGCTGGGACCAACATGATTC	291
Qy	242	TTTTCAATCAAGTTTTGGCTTTGTTCCCTGTGCTATTATCTTATATGCTTATTTATGCAAA	301
Db	292	CTTCCCTTAAGTTTTGGAAATGTGCCGGAGGCTGTTATTTGTAACATATGCTCATGGGTG	351
Qy	302	GCTAAGTATATTTTAAAGTAAGCAGGCCCTTTATTTATGACAGTGGGAACGGCCTTTTAAAT	361
Db	352	GTTAGGACATCGGTATCTCGGGGATACCGGTTTTTATTTGCTTCAATGGCCCATCTCTGG	411
Qy	362	TTTTCTTGGCCGTGTCCCGACTGTAATTTATCCGCTACGGAGTGTTTTACATCCTACAGA	421
Db	412	TTTTTCTTCTGTTTGGTGTGATCATCATTTATCTGATGGGGAAAGCCTCATCTCAACTC	471
Qy	422	ATTTCGTGACGGTTTACAGGGCATCTCAACCCGAGATTTGGATGGAGCTGCTGCCATCTT	481
Db	472	TTCCCTGATTAATTTACAGAGCTCTCTCTCAAGGACTTGTGGTTTTTATGTATGGT	531
Qy	482	AAGAAACTGGACATTTGCTGCATTTTATGTATCTTGTAACATATGGGGAAGCCTCATGCT	541
Db	532	CCGTTACTGGAGATTACAGATTTTATACGTATATGTACAGACTGTGGAGTTCCGTTGCTCT	591
Qy	542	ATCTCTATGTTCTTGGGAGTTTGGTAAATTAACAATAATCCAGAGCAAAAGCGTTT	601
Db	592	TTTCGATGTGTTTGGGGACATAGGCCAATCAATATACATATTAACGAACGGGCGGTTT	651
Qy	602	CTACCCCTTTTTCGGTATCGAGCATTAATTTCTTTCTAGTCTTGCTGCTGCATATTTG	661
Db	652	TTAGCCTCTTATCAATACAGGATTAATCTCTCTCATATTTGGCAGAGAAATCTCCCTA	711
Qy	662	TTGGCCTTCAAGTTTGAGAGCTTCCGTTTCTGAAGGTGTAGATCTTGGGGAAATTTCTTT	721
Db	712	TTGGATGGGGAACAACATTTGTTGGCTACTCCTTTGCATGTGATTTCCGGCACCTGT	771
Qy	722	ACGTTTTTGAATGGCTATGAGCATTTATGATCTGACCTGTCTATATGGCAGTTACTGGTG	781
Db	772	AATGCTCACTTGAACATGCTGATCACTGTTGTTGTTTAAATTAATCTGGCTAATATG	831
Qy	782	GATCAATAGAAGCTATTGACCGATCTCGCTTC-----TA	817
Db	832	GCGGATTCATCATTTGATTTGATTTACTTGTGATGCCCTCCATCTAGACGTTGCTGGCAGA	891
Qy	818	TATATCCACAACAATGCCAAAAGGGGAAAAAGGTGCTAAACCTTAATAATATGTGAAGA	877
Db	892	AGAGGACACACTCTGCTAATCTTAAAGGAAAAAGAAAAACCTTAAAGCCAAAGCTAGAAA	951
Qy	878	TAGCTTCCCTATCTATGTATAGATTCCTTATATTTCTTTATTTAACTCTCTTGCTATTGC	937
Db	952	CCTTTTCTTACACCTCATATGCTGTCTGTTATTTAATTAAGGCTGCTAATTAATGTCTATC	1011
Qy	938	CTATGCTATTGTCATTAACTTAATTCGAAGTGACTTGGAAAAATGACACTGAACTGCATA	997
Db	1012	CTATAATTTGGTGTATCATCATCTATTCCAGAGTCGTTTGGAAAGATCAAGTTAAGCCAGATT	1071
Qy	998	TCTCAATATGAATACATATATAGTGAAGTTCATGAGGAACCTTCTTCTGGAAGTGGCTACT	1057
Db	1072	CAGTTCTCAGCTAACAATTCATATGGGTATATAGATAGAACTACATACCTCATTTGGGTGCT	1133
Qy	1058	ATCCGTAATTATCATATGCTATTTGTTGGTGTAAAGTCATATTCGTAAATTTGGATGTTAAC	1117
Db	1132	TTCTGTATTTACACCTGTACCTCTTACCGAGACGTGATATCCGTAAATTTGGGATGTACGT	1199
Qy	1118	TGGAGCCCTATGCTACTCTGTGATGGTTCCTCCATAACAGATATGCTTTTCTTGCTCTTGT	1177
Db	1192	CGGTGCTTTTACTCATCTCATTTGGTAATAGTTTACAGAGACGTCTCTTTTGGGAATAT	1257
Qy	1178	TATCTTTTGAAGAACCAAGCT---TCTGGGCTGGTGGCATGTGTCGGTATGCGTAACATCTCTAT	1234
Db	1252	TTTTTGTGCAAAAAGACATCTCTATTTTGGGGGAGTTCTTGGATATACACCTCTGGC	1311
Qy	1235	GCTAGCTGTGTGTTGTCGGAGCTATATACAGATTTATTTTGGAAATCCACAAATATAGCTCT	1294

Db 1312 TCTAGCTGCTGAGCTGGAGGATGCAAAATGCTCTATCCCGGGGAGCAAAATTACGTT 1371
QY 1295 CTTTACGTCACACTAAGAAGTGGCCATATCCCTCTTGACCAAGCAAAAGTCAANG 1354
Db 1372 CTTTATCAAAACCAAGAAATGGCTTTATCCACTTCTTCAGAGGATAAATATCANG 1431
QY 1355 TAAGCTCTATTGATGATGAGTGGCCGCTTGGAAAATCAGAGAGGCTTTATCA 1414
Db 1422 GAACCCCGGATGATGATGCTGTTCAAGGATAGAAAGTCTGGAGGCTCTTTATTTA 1491
QY 1415 ACAAGTTTGGCTGTTATCTGTGAAGTATGAGCATGACCCCTTATCTTGCAGTAT 1474
Db 1492 CCAAGGCTGCTGTTATTTCTCTTCCTGACGCAAGTTTAAAGTCATCGCCCTAT 1551
QY 1475 TCTCTTTTTCATCTTGCATTTGTTGGTTTCTCAACTAGTTAAACAATCTTCTT 1534
Db 1552 TCTCTCATTTATGATGCTGTTGATGCGGTTGTTGCTTATTCGTAAGAAATACTA 1611
QY 1535 AGCGAGTCTGCTTTAAAGAACAGAGTGGCTCAGAGATTCAGCTCTGCTTC 1591
Db 1612 CTTAGAGCTGCTGATGCTGATGCAACCTTGAACACCTAAAGAACTTCTCTTC 1668

RESULT 19
AE001646 10163 bp DNA linear BCT 01-DEC-2000
LOCUS AE001646/c
DEFINITION Chlamydia pneumoniae section 62 of 103 of the complete genome.
ACCESSION AE001646 AE001363
VERSION AE001646.1 GI:4376910
KEYWORDS
SOURCE
ORGANISM Chlamydia pneumoniae CWL029.
REFERENCE 1 (bases 1 to 10163)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 2 (bases 1 to 10163)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source location/Qualifiers
1. 10163
/organism="Chlamydia pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
complement(93..1715)
/gene="adt_2"
/note="Cpn0614"
complement(93..1715)
/gene="adt_2"
complement(93..1715)
/codon_start=1
/transl_table=11
/product="ADP/ATP translocase"
/protein_id="AAD18753.1"
/db_xref="GI:4376911"

gene
complement(1944..2450)
/gene="pgsa_1"

gene
complement(1944..2450)
/gene="pgsa_1"

CDS
/note="Cpn0615"
complement(1944..2450)
/gene="pgsa_1"
/codon_start=1
/transl_table=11
/product="Glycerol-3-P Phosphatidytransferase"
/protein_id="AAD18754.1"
/db_xref="GI:4376912"
translation="MRPQNLSLRMLALYFCOEKIHILLAIYGNMSDYDGL
ARRKATSLGSLIDPTDKVFEVCTIVLYMESLSIAHLFFICARDLFLIFCYL
SLVGMKGVDYGLFWGKIFTVVOFILLGVTAGGEIPMTGLVPLVGLFLEIRIM
DYKKQFLR"
2672..2753
/product="tRNA-Ser"
3104..4453
/gene="dnab"
/note="Cpn0616"
3104..4453
/gene="dnab"
/codon_start=1
/transl_table=11
/product="Replicative DNA Helicase"
/protein_id="AAD18755.1"
/db_xref="GI:4376913"
translation="MIVGCMITGVHYLNLAAANDYEEDEFYLEHKIIFRYLDARFQ
DKPIDVHLAGEELKRHNQITVIGGPSYLTITAEFAGTAAYLEEVVDIIRSKIM
STAKIERALAEOKRNAVALEDEMONSFRTISQTSVSQTLVADKLRLGTLTTDPK
LVQDERELFLNAGDNKSFETGIPHTIDLDLIHGFSPSNMLTAAAPAMKTA
LALNIAENLCFONRLPIGIFSLFEMTVDLIHRMICSREVSQKISTIDLSGHFOR
VSVINEMOEHLLIDDPGLKVSDLRARARMKESYDIOFLIDYLDLSSGSLRAT
ESRQTEISEISRMKTLARELINPILCLOSRLKVEDRANRPMMSDLRESGTEODS
DLVMPLEIREYVDDNDKGTNELIIAKNRHSISVPLVFEKELARFNYSAPEDIS"
4797..6632
/gene="g1da"
/note="Cpn0617"
4797..6632
/gene="g1da"
/codon_start=1
/transl_table=11
/product="FAD-dependent oxidoreductase"
/protein_id="AAD18756.1"
/db_xref="GI:4376914"
translation="MMTRPIAVDYIVVAGHAGEAAYCSAKMGVSLMLNSLDTIA
KISCPAVGIGKGIYREIDALGIMAEVVDOSGIORILNORKGPAVRAPRQVOK
OLYTHMKRLLENPGHIMATYBSLDKRGYISGYTTKGMFSGVTVLSSGTPM
RGLIHIDGRNFGRLGDPSSQGSSEDLKRGFTLSRKTGTPRLASSINFCMEB
QPGDVGVEVHRTPEFQPLPQLSCFTITMEKTAIISAMLRSAALYGCITFEGPR
YCPSIDKIVKSDKERHVFLEPGLTOEIVANGISTMPPFOVDMIRSVGLGN
AITRPAIAYDYIIGNVIHPTLESKLEGLFCGQINGTGYEAAAGLIGLINA
VNKVENRPFIPSRQESYIGVLDLDTQIOLDEPFRMPTGRAERHLRLRODNACARLS
HYGEIGLSEERYEYKONOLLEEVRYROKFEROVGGSVSLARLSRPVSYDM
LREAFPNDRICAVLANSLEMEIKYSYIDROKILIOSLEKASLSLIPEDLDYKQIT
ALSLAEQEKLAFTPRYLGSASRISGINSADIOYMLALKRHHH"
6619..7326
/gene="ip1a_2"
/note="Cpn0618"
6619..7326
/gene="ip1a_2"
/codon_start=1
/transl_table=11
/product="Lipote-Protein Ligase A"
/protein_id="AAD18757.1"
/db_xref="GI:4376915"

gene
complement(7323..7757)
/gene="ndk"
/note="Cpn0619"
complement(7323..7757)
/gene="ndk"

CDS
complement(7323..7757)
/gene="ndk"

Oy 1295 CTTTGACTCACTAAGAAATGCGCTATATCCCTCTTGACCAAGACAAGTCAAG 1354
||||| ||| ||||| ||||| ||| ||| ||| |||
Db 444 CTTTGATCAACCAAGAAATGCGCTTATCCCTCTTGACCAAGAAATCAATG 385
Oy 1355 TAAGCTGCTATGATGATGAGTTGCGCGCGCTTGAGAAATCAGAGAGCATTTAATCCA 1414
||| ||| ||||| ||| ||| ||| ||||| ||||| |||
Db 384 GAAACCCCGATGATGCGTCTGTTCAAGATAGAAAGTGTGAGCGCTTTAATTTA 325
Oy 1415 ACAAGCTTTCGCTATCTGAGAGATTTGGAGCTATGACCCCTTATCTGACAGTAT 1474
||||| ||||| ||||| ||||| ||||| ||||| |||
Db 324 CCAGAGGCTGCTGATTTCTCTTCTGTCACAGATTTAAACGTATCGCCCTAGT 265
Oy 1475 TCTTCTTTCATCATTTGATTTGCTGTTCTGCAACTAGTAAGTAAACAATCTATTCT 1534
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 TCTTCTCATTTATGATGCTGTTGATTCGGTTGTGCTATATCGTAAGAAATACGA 205
Oy 1535 AGCCAGTCTGCTTTTAAAGACAAGAGTGGCTCAAGAGATTGACCTCTGCTTC 1591
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 CTCTAGAGCTGCTGATGCTGATGAGCAACCTTGAACACCTTAAGAACCTTCTCTTC 148

RESULT 20
AE002174 10407 bp DNA linear BCF 30-MAY-2000
LOCUS Chlamydia pneumoniae AR39, section 10 of 94 of the complete
DEFINITION genome.
ACCESSION AE002174 AE002161
VERSION AE002174.2 GI:8163367
KEYWORDS
SOURCE Chlamydia pneumoniae AR39.
ORGANISM Chlamydia pneumoniae AR39.
REFERENCE 1 (bases 1 to 10407)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935
REFERENCE 2 (bases 1 to 10407)
AUTHORS White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Jun 1, 2000 this sequence version replaced gi:7189059.
FEATURES
source
1..10407
/organism="Chlamydia pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
124..206
/gene="rRNA-Leu-2"
124..206
/gene="rRNA-Leu-2"
/product="rRNA-Leu"
343..1305
/gene="CP0125"
343..1305
/gene="CP0125"
343..1305
/note="conserved hypothetical protein; identified by
glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

/protein_id="AAF38008.1"
/db_xref="GI:7189060"
/translation="MRLSILKLHLFLSRSSSSLSPHYHSCSPSMHLILCRMKDADI
MEMQDITILSGVCSRMGKLVLSQETQDSCHQDEHERHILQIYEQULSALDEETRRRE
EAKNODLEKLOQENTWLONRLEAKLQOI RHQSDI IDETKRELDSVQETISEGRRLC
YEHIKOLEBOLQRYVSOHGAPEISIEDEKSSAAVAELNRLKSLIDLOQEKDYIKT
YHSELAKEKLEKLOREQOTSSEVCSIEKLEVEQTDLAERKKAIALLODIVEDQYCOL
RLHKEMGAMPSTMKLDHLKGLGKEPESVDVVFSEKSLGS"
1448..1954
/gene="CP0126"
1448..1954
/gene="CP0126"
/note="similar to GP:999741; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="Holliday junction resolvase"
/protein_id="AAF38009.1"
/db_xref="GI:7189061"
/translation="MSELIGVDPTIVAGYAITAVEQRYOLRPYSGARIRLSDMPL
PMRYKTLPEQLSGVLDPTOPNANVLEFQVFNKPSQSTKLMARGIVLLAAQORDILI
EYAPNNAKNAVVGKGNASKROVOVMYSKILNPEVLHPSNEDIADAFALAICTHVA
RSLQGVK"
1956..2579
/gene="CP0127"
1956..2579
/gene="CP0127"
/note="similar to PID:1183841 SP:Q51425; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Holliday junction DNA helicase RuvA"
/protein_id="AAF38010.1"
/db_xref="GI:7189062"
/translation="MYDIKQTLVYHTGAIVIECGIGYHAIATERNAIECTIRALHQ
DLVFLVTHIFRETEHLVGFHSREBERCFRLISFGIGPYLALAILNALPLKVCV
VRSIEDIRALAVSGIGKTAELKWLKOKPLDLSRVTSTHTSSCLEEGIQ
ALAAVGYKIAERMIABEAIKDLPEGSSLTDLPLAKKNSGVNKD"
2651..3085
/gene="CP0128"
2651..3085
/gene="CP0128"
/note="similar to SP:P2423 GB:X57555 PID:416172 GB:U00096
PID:1788866; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAF38011.1"
/db_xref="GI:7189063"
/translation="MEORTLIIKPDPSVSKAHIGETLSIFEOGSLRIAMKMMHLSOTE
AEGFVHREBPQELVDPMVSGPVVVLVEGANAVERNREIMCATNPAPASGTR
AKFGESTVNAVHSGSDTLNNAVEIAIFFSKIEVYNAKPLV"
3085..3789
/gene="CP0129"
3085..3789
/gene="CP0129"
/note="similar to GB:U14003 SP:P32099 GB:X03046 PID:432634
PID:504496; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="lipote-protein ligase-related protein"
/protein_id="AAF38012.1"
/db_xref="GI:7189064"
/translation="MPTTNCITFLDRGSHILHLOLIEALLRVANONCTINSKAKDS
IVLGISRLNDVHISRAQADHPIPIIRYSGGVTFIDSNLWVSWINNSSEASQPO
ELAWTVGIYSPLNPFSIRENDVGHKRGNAQYIQHRVWHHTTFLMADIDLK
LSYIPLIPQOQPTVNRQSHSEELFTLRPMPSPRDLERIKASGLFTWEEFLDNE
LEEIIAOPHRAKATVNLN"
3789..5611
/gene="CP0130"
3789..5611
/gene="CP0130"
/note="similar to GB:L10328 SP:PI7112 GB:K00826 GB:X01631

OY	938	CTATGCGATTMTTGCAATTAACCTAATAAGAGTGCATGGAAAGGTACGAGTGAACATGCAATA	997
Dd	9604	CTATATATTGGTGATGCCACTCTATTTCCGAAGTCGTTTGGAAAGATCAAGTTAGCCAAATTTA	9663
OY	998	TCTTAATATGAATGACATATAGTAGAGTTCATGAGGCAACCTTCCTCTTGAGCTGGCGTAGT	1057
Dd	9664	CAGTTCTCAGCTAGAATTCATAGGTTATATAGTAGATCATCTACCTACCATTGGGCGTCT	9723
OY	1058	ATCCGCTACTTATCATGCTATTTTGTGTGGTGAACGTCATTCGTAAATTTGGATGGTTAAC	1117
Dd	9724	TTCTGTATTAGACGCTGACTCTTCACCGGACAGTGTATCCGTAATATGGGATGAGACTGT	9783
OY	1118	TGGAGCCCTAGCACTCCTGTCANCGTCTCCTAACAGGTAATCGTTTCTTCGCTCTGT	1177
Dd	9784	CGGTGCTTTGATCACCTCATTTGGTAATATTAGTTTCAGAGACTGCTCTTTTTCGGAACAT	9843
OY	1178	TATCTTTTGAAGAACCAAGCT---TCTGGGCTGGTGCCTATGTTTCGGTATACACTCCCTCAT	1234
Dd	9844	TTTTGCTGCAAAAAGAACATCTCATATTTTGGGGAACTTCTTGGAAATGACACCTCTGGC	9903
OY	1235	GCTAGCTGTGGTGTGSGAGCTATACAGAAATATCTTTTGCAGAAATCCACAATAATACGCTCT	1294
Dd	9904	TCTAGCTGCTCGAGCATGGAGGATGCAAAATATGCTCTATCCGGGGAGCAAAATTTTACGTT	9963
OY	1295	CTTTGACCTACACTAAGAAATATGGCCTATATCCCTCTTGACCAAGGACAAAAAGTCAAAGG	1354
Dd	9964	CTTTGATCAAAACCAAGAAATATGGCCTTTATCCCATCTTCTCCAGAGGATTAATAATCATGG	10023
OY	1355	TAAAGCTGCTATATGATGATGATGCGCCGCTCGGAAATACAGAGAGAGCTTTATATCCA	1414
Dd	10024	GAAGCCCGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	10083
OY	1415	ACAAGTTTGGCTCGTTATCTGTGAAGATATGGAGCTATGACCCCTTATCTTGCAGTAGAT	1474
Dd	10084	CCAAGGCTGCTGTTGATTTCTCTCTGTTGACACAGATTTAAACGTCATCGCCCTAGT	10143
OY	1475	TCTTCTTTTCAATCATGCTATTTGGTGTGTTTCTGCAACTAAGTAAACAAACTATTTCTT	1534
Dd	10144	TCTTCTCAATTTATTAAGTCGTTTGGATGTCGGTGTGCTCTATATTCGGTGAAGAATPACTA	10203
OY	1535	AGCGCAGTCTGCTCTTAAAGAACCAAGATGGGCTCAACAAGTTTCAGTCCGCTGC	1591
Dd	10204	CTCTAGAGCTGCTGCTGCTGTAAGCACTTGAACAACCTTAAGAACCTTCTCTCTTC	10260
RESULT 21			
AP002547/c			
LOCUS	300550 bp	DNA	linear BCF 25-MAY-2002
DEFINITION	Chlamydia pneumoniae J138 genomic DNA, complete sequence,		
ACCESSION	AB002547		
VERSION	AB002547.2		
KEYWORDS	GI:10176693		
SOURCE	Chlamydia pneumoniae J138 (strain:J138) DNA.		
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
REFERENCE	1		
AUTHORS	Shirai,M., Hiraoka,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M., Takeuchi,A., Nishida,J., Shibata,K., Fujisaga,R., Yoneda,H., Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A., Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T. Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of Chlamydia pneumoniae isolates from Japan and the United States		
TITLE	J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)		
JOURNAL			
MEDLINE	20298986		
REFERENCES			
AUTHORS	Shirai,M., Hiraoka,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K., Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T. Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA		
TITLE	Nucleic Acids Res. 28 (12), 2311-2314 (2000)		
JOURNAL			
MEDLINE	20303439		

REFERENCE	AUTHORS	TITLE
3 (bases 1 to 300550)	Shirai,M.	Direct Submission Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail:mshirai@pe.cc.yamaguchi-u.ac.jp, Tel :01-836-22-2227, Fax:01-836-22-2415) On or before Sep 15, 2000 this sequence version replaced gi:61722398, gi:6172300, gi:61723396, gi:61723398, gi:6978889. Location/Qualifiers 1..300550 /organism="Chlamydomonas reinhardtii J138" /strain="J138" /db_xref="taxon:138677" /note="synonym: Chlamydia pneumoniae (strain J138)" complement(21..794) /gene="dapB" complement(21..794) /gene="dapP" codon_start=1 /transl_table=1 /product="glutamate epimerase" /protein_id="BAAB8725.1" /db_xref="GI:8978890" /translation="MAPYSPSTSKRYFTYSAGNRFLLGELTPEVEDRELCQETRVDFGLKPPSCADQLITFNSDGRPTMGNGLRCAIAHLASOKKSDISYSTDSGLYSGYFSMDRLVLDMTLADMRAVSVRLESRPDLPEVCCHTGVAHVILPEISTDLSTLPFLRYHQTFSPDGYNVFQILGHCOLRVRTYERGVGETAAGCTGAALAVVSNSSCKKESTDIHWGGEELMVSNRRNVYIQGSVTEDL" complement(763..1338) /gene="clpP_1" complement(763..1338) /gene="clpP_1" codon_start=1 /transl_table=1 /product="Clp protease" /protein_id="BAA98726.1" /db_xref="GI:8978891" /translation="MADGEVHKLRDIIIEKLLEARVFEPSPTYESASDAIKKWLWELRKGRPIRVINSRGSGVACRANVDQIKMLTPSYTVTLGASMSYLSICAPGRRTPHSRIMIHQPSIGCFITGOATDDLIDHAEIILKTARIIDIVEATNOPRDILLKALDRDMVMYANAKPFGLDIGLFESFDL" complement(1358..2917) /gene="glaA" complement(1358..2917) /gene="glaA" codon_start=1 /transl_table=1 /product="serine hydroxymethyltransferase" /protein_id="BAA98727.1" /db_xref="GI:8978892" /translation="MKVEKKFKFAIVETTKVVAVYSLHKLFLNASKKGKOSTASTAYIALDLHLNAFPSTIGERTIIDLKRSRSHLKAISENNSSLSVOAMGNLTDKYCEGSPKRRFYSCCENVDAILMEWCETLARLEPADACVQPHSGADANLLAWAILTHRVQGPVASKYGTVELTEEBEYTLTKAESMSCVCGPSSLNGLTHONVRLNWSKLMTRCFYDVNPDEDECDDYEISRLEAKYPRKVLIAYSYSIRNLNFVAJOAEDGCVLVWDNAHEGLTAGVGVPDEENRPYADIIVTTTKTLRGPRGGVLANDREESTLNKAKCPIMGGRPLPIVIAKTYALKREALSVDFKRAHOVYVNARRLARELSHGRIILTGTDNHMMVDELGSLGSKIADIEDILSSVGIAVNRNLPSDAIGKWDTSGIRGTPALTLLGMGIDMEEVADVIIVKLRINIRLSCHVEGSSKNKGELPEAIQAEDARNLRLRPPLYPEIDLEALY" 1107..3378 /gene="CPj0522" 3107..3378 /gene="CPj0522" codon_start=1 /transl_table=1 /product="CT433 hypothetical protein" /protein_id="BAA98728.1" /db_xref="GI:8978893" /translation="WTLYLGINKQTARKYQAHYLPILTPEPAKSTPONKRALQFLPDATHVILSPSSTHFLSRNMSTLSKATLKTKTYTCIGESTKERLLSFAGOVKVVAQO


```

AFSVSIFQNPQILLDEVEAAGDSYFEKSSLNLMKNEKFNPIISIVSHOEELIKDNC
DRCILKGDHGIIDGSTPSEIFRIYKQOSNKEIHK"
gene
2064..2840
/gene="RP004"
2064..2840
/gene="RP004"
/codon_start=1
/transl_table=11
/product="O-ANTIGEN EXPORT SYSTEM PERMEASE PROTEIN RFBA
(rfbA)"
/protein_id="CA14477.1"
/db_xref="GI:3860576"
/transl_table="MIKYEFSKKYWRATALLKVASIVRONKDSFGLSMLSIOPFIH
IVISYFEGGLTOPREFEWMNVIGIPLMSFVSSLTICASLWTRDOIKKRVSKT
FEPIADSGOLLYTLICFSFAMNFAFILLEPEKFSQIIEPILILPIVICYSSIAV
AFIPIYRIDIPOMLVILIGVIMSPIDYIPYSLIPESKRIYFEENPFVLIRPVQVL
INGSLPDMILIKSVIAFIVFISYLIYRQPSKAVITVL"
2837..3724
/gene="RP005"
2837..3724
/gene="RP005"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14478.1"
/db_xref="GI:3860577"
/transl_table="MKLGPNLDFNELDITGLKILDOIIFDLVLRKADSLHKDMLFRA
TPLSIIPODYSKFLKISPHLDDPLAEFLCISKEVTIRLKHKDDIITYECKRKVOR
VAVKYPLEKIKIDIDEDVYKLTMLIGTNFTSRFAQIYIWOODESFSLIEDIAA
RYAARVFSYVNSLSNMKNLFLQNDILFNLQOKLDKKNLIDDKILRYOKNERVDFD
YKDSFENLDEALNNSHCICYCHKRDKSCSGFDVYIPHRDVIYSVKILPRLSYRMTG
CPLKOKISEMNYIKAOVLI"
3757..5877
/gene="RP006"
3757..5877
/gene="RP006"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14479.1"
/db_xref="GI:3860578"
/transl_table="WVAATGHRICDCKSAKYOKODEPVNPILESNILEETLKLPG
LEIYILLTRMPLNIYAPLPEPTNYNLTVLGAPAGSLAYLLRSHNVTAIADGLK
ITPLSFNTHKPYKFWHEKKNLSEIRIIPKFGVGAEGYITIRMDKKNLIDILIRRN
NKYYDVALGPNITRBOALDLSFDHVAFCJGAGCKPKILNINENFAKGVRTASDPLM
LONGAFIQNSNTNMIMRPIYITIGGITSIDVATESLPIYKKOYEEPAKNTIEOILT
BEDKIAIEFTAHAKLEKANEKNEELKVFNLGATVYRGRLODSPAAYKLNHELI
YTLAIVNFKEMOPLRINIDKYGHVESVESITMLDRTHKKYHNMCPYKPSN
TVLKTQTVMAIGIENNTQEDYDYKSYGDCNPKYFGSVYVAITSAREGEYVIRKL
IDNVSFKGSYACFTLODYLITSRINKINILNDKTFELIHSPLAKNFOGOEFLR
ONYSKDITKLEPVALSPADIDIEKGLISFTIYEVGKSTSLCKTISENKKVYLMGPTG
SPLEIPONKRTIIIDSKYRNNGLLKILEKNKRVIFATYPPDIKKHKLISVDIYIINTS
PEIAEIDELKIFGENTELLINVNSLMOCMMKMGICGCIQKVKOKYIFACSEQON
VEIDFKSLKRLRONSLOEKMS"
gene
complement(6227..7021)
/gene="RP007"
complement(6227..7021)
/gene="RP007"
/codon_start=1
/transl_table=11
/product="ACYL-[ACYL-CARRIER-PROTEIN]-UDP-N-
ACETYLGUCOSAMINE O-ACYLTRANSFERASE (lpxA)"
/protein_id="CA14480.1"
/db_xref="GI:3860579"
/transl_table="WNSNINHTTAIIAAGANFGKNVQPYCIIIGPEVVLHNDNELKS
HVIDIGITEIGENTVYIPFASIGPPQILKANEKSSYIIGSNNTIRREYVVOAGSKS
GGMITRVGNNNLFFMVGHIGHCKIGNNLVEFANYSLAGHAKVGDYAIIGISAVHOY
TRIGEYSMIGSLPGADVIPGGLVSSRAVEGNLIGMNRKGGKADSLTALNAVE
EPIEGENPVDRIKOVAREKKNNSIVTQIDIFLNDSSRACHFEK"
gene
complement(7031..7468)
/gene="RP008"
complement(7031..7468)
/gene="RP008"

```

```

/codon_start=1
/transl_table=11
/product="(3R)-HYDROXYMYRISTOYL-[ACYL CARRIER PROTEIN]
DEHYDRATASE (fab2)"
/protein_id="CA14481.1"
/db_xref="GI:3860580"
/transl_table="MTAITTEIMDLIPHRYPFLVDVRLKIDPNKSLIGIKNVNPE
OETGHPARVPWPGVLWESMAQALAILVAKSIDSTKKEVFLMSIENTKRRRIYQPS
DTMHHSVLDQORAWMKFESSKVMYECIELASKRTAKIKMK"
complement(7659..8699)
/gene="RP009"
complement(7659..8699)
/gene="RP009"
/codon_start=1
/transl_table=11
/product="UDP-3-O-(3-HYDROXYMYRISTOYL) GLUCOSAMINE
N-ACYLTRANSFERASE (lpxD)"
/protein_id="CA14482.1"
/db_xref="GI:3860581"
/transl_table="WVSNFYNKNGIPRRLTAIVDFLHDIIETPKYEDIDYDIKILQ
EASPDISFLSPKISEFLKTTKAACVPAKNTFEVNQNTYLLHAENSYRASKLID
FETAPTKSYSTRIMKSAIIDSATIGKNCYIGHNVIEDDVYIGDNLIDAGFTIGRG
VNIKNARIQHVHSINVAIIGDDVVLVGARIGDGFSETEKGVNHRKIFHIGIVKIG
NVEIGSNMTTIDRGALODTIIEDLCRIDNLVOIGHGVKIGSIIYAOAGIAGSSAIG
KYCALGQGVAGIAGHLNIGDGTQVAAQGVAAQNEBCKIVGCSPAVPIQDMHROSITMK
OLVKSNSKIKK"
complement(9267..9339)
/gene="tRNA Phe (GAA)"
complement(9267..9339)
/gene="tRNA Phe (GAA)"
9661..10666
/gene="RP011"
9661..10666
/gene="RP011"
/codon_start=1
/transl_table=11
/product="NIFR3-LIKE PROTEIN (nifR3)"
/protein_id="CA14483.1"
/db_xref="GI:3860582"
/transl_table="WLVNDIMLNPITFIKIGNIELSSNVILADMSVDLTLEFKLV
KRFAGLAVNEMIASRAMIKMSROSOMCAIMHDPISACVOLACCEBDAEAAKMN
EDMGAKIIDLNEGCPARKYGVGASALMBEOLAKIFEATVKAVKIPVYKMRIGN
DONTNAPNTEKLIANSGVQVTVYGRTRCQPSGNANMDFRTYKBAVKIPIVLANGD
ITNFNAKEALIDORSAGDIMSIVGRGVYKGPWLISOIAYLTKGKEKAPASIAEOIIT
KHYDIDYVYKGSVSPYARKHIIWYNSGLPSSAFECTVNLMDPIAYKEKIAEFTM
SVMDANK"
10683..11756
/gene="RP012"
10683..11756
/gene="RP012"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14484.1"
/db_xref="GI:3860583"
/transl_table="MKRILSFIFILFENSSAYAKAAIITDYKPIFLPUTENKKIK
IAIRSYLKNKKSYPVLDVDPNSEKTEIVLOELVILPANKIEKNLILKLNKSYIKALN
KYNFDOSLOSFKODEFSNKFPTDNTIYVEHKLNKSSHSSTFMNSTVOANYGARSSMY
Query Match 17.9% Score 293.2; DB 1; Length 282610;
Best Local Similarity 53.1%; Pred. No. 1.2e-59;
Matches 781; Conservative 0; Mismatches 653; Indels 36; Gaps 6;

```

Db 59434 TCATTTAAAGACGCTTTTGTAGTACAGATATAGTA-----CAGATCGATAGTTTT 59487

QY 246 ATCAAGTTTGGCTGTGTGTCCTCCCTGCTATTAATCTTATGCTTATTTATGCAAGCTA 305

Db 59488 TTAATAACATATATAGTACTACCTTCTGCTAATTCGTATGATATATTTATGTAGCTA 59547

QY 306 AGTAAATATTTTAAAGTACAGAGCCCTTATTTTATGACAGTGGGAAGCCCTTTTAAATTTTC 365

Db 59548 TGTGATATTTTAAAGCAAGCAAAACGTAATTTATGTTATATCTTATCTTATTTTAAAGGAT 59607

QY 366 TTTGGCCCTTCCCGACGTAATTTATCGGCTACAGGATTTTTCATCCCTACAGATTT 425

Db 59608 TTTGGCTTATTTGCTTTCTTCTTCTTACCAATATCTGATTTAGTCCAGCTGATCTAA 59667

QY 426 GCTGACCGTTTACAGGACCTCCACCTACCTCAGATTTGCTGAGATCGTGTCCATCTTAAGA 485

Db 59668 ACTATAGATCTTTAGTTTACCTTATCTTAATTTCAATTTGTTTATTAATAATAGTTGCT 59727

QY 486 AACTGACATTTTGTGCTGATTTTATGCTGCTGCAACTATGGGGAAGCTCATGCTATCT 545

Db 59728 AAATGAGATTTTGTGCTTCTTTTATACATATTCGCGACCTTTGGGGAACAATGATGCTTACT 59787

QY 546 CTAAATGCTGGGATTTTCTTAATGAATTAACAAAATCCAGAAAGCAAGGCTTTCTAC 605

Db 59788 TTAATATTTTGGCAATTTGCTAATCAATTAATTAATGCTGAGAGCTTAAGGCTTTCTAC 59847

QY 606 GCTCTTTTGGTATCGAGCTAATATTTCTTACTAGCTTCTGCTGCTGCAATTTGTTTG 665

Db 59848 TCAATGTTTGGTTTACTTCCGAAATTTAGCATTCGCTGATCAATTCAGTGTATTTGATAT 59907

QY 666 GCTTCAAGTTAGAGCTTCCGTTTCTGAGGTGATGATCTTTGGGGAATTTCTTACGT 725

Db 59908 TTTTCAACGAAAAAAGCT-----CAATAAGTTCGCAACATTTAAATTTGTA 59955

QY 726 CTTTGTATGCTATGCTATGATCTGATGCTGCTTCTTATGCGCAGTACTAGTGTGATC 785

Db 59956 CCTTATTTGTTATATGATATGATACAAAGTATTTCTTATATATTTACATATGATGAGAT 60015

QY 786 AATAAGAACGTATGACCGATCTCGCTTCTATTAATCCAGAAAGAAATGCAAAAGGGGAAA 845

Db 60016 AATAAATATGTTCTAATGATGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 60069

QY 846 AAGAGTGTAAACCTAAATGATATGAAAGTATGAAAGTATGCTTCTTATGATGATCTCT 905

Db 60070 AAAAAAAGCTAAAGCTAAATGCTGTTCAAGAAAGCTTTAAATGATCTTTTACTTCGAAA 60129

QY 906 TATATCTTTTATTTAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 965

Db 60130 TATGATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60189

QY 966 GTGACTTGGAAAGTCAAGCTGAACTGCAATTCCTAATATGAAATGACTATAGTAGTTTC 1025

Db 60190 GGTGTTTGGAAATCCAAAGTAAGAAATATATCCGACAAAGAGGCTTATACCAATATAT 60249

QY 1026 ATGGGGAATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1085

Db 60250 ATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60306

QY 1086 GGTATGCTATCTGTAATTTGATGCTTAACTGAGCCCTGATGCTGCTGCTGCTGCTGCTGCT 1145

Db 60307 AGTATATTTTAAAGAAAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60366

QY 1146 CTCCTAAGAGTATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1205

Db 60367 TTTCAATAGAGTGGCGCAATTTTTCATTTATATTTTTCATTTGATAGGCTTATGCAATGA 60426

QY 1206 GT---CGTATGTTGGTACAACTCTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1262

Db 60427 TTAACCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60486

QY 1263 AATATTTTGGAAATCCAAAGATAGCTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1322

Db 60487 AATGTTTAAAGTAAAGTGTGTAATATTTCTTATTTGATGCACTTAAATATATATGCGCTAT 60546

QY 1323 ATCCCTCTTGACCAAGCAAGCAAAAGTCAAAAGTAAAGCTGCTATTTGATGATGCTCCGCC 1382

Db 60547 ATTTCCACTTGATAGAGATTTTACAGAGCAAGGCAAGCTCCGCTTGAAATTCGAGAGA 60606

QY 1383 CGCTTGGCAAAATCAGAGAGCTTTTATTCACAGAGCTTTGCTGCTATCTGGAAGT 1442

Db 60607 AGCTCGGTAAATCAGAGCGCTCTATTTATCAATCTTATTTATTTTATTTTATTTTATTT 60666

QY 1443 ATTTGAGCTAT-----GACCCCTATCTTGAGATGCTTCTTTCATCATCTTCTGCTAT 1496

Db 60667 TTTGCTTTTATAGAGCGCTCTTATTTTCTTCTTATTTCTTATTTATTTATTTATTTATTT 60726

QY 1497 TGGTGTGTTTCTGCAACTAAGTTTAAACAA 1526

Db 60727 TGGATATTTGCAAGTTTAAAGTTTAAATATA 60756

RESULT 24

AE008575 14117 bp DNA linear BCT 14-SEP-2001

LOCUS AE008575

DEFINITION Rickettsia conorii Malish 7, section 7 of 114 of the complete genome.

ACCESSION AE008575 AE006914

VERSION AE008575.1 GI:15619111

KEYWORDS

SOURCE

ORGANISM

Rickettsia conorii.

Rickettsia conorii.

Bacteria; proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.

1 (sites)

Ogata, H., Audic, S., Barbe, V., Arcligneave, F., Fournier, P.-E., Raoult, D. and Claverie, J.-M.

Selfish DNA in protein-coding genes of Rickettsia

MECLINE 20485642

PUBMED 11030655

REFERENCE

2 (sites)

Ogata, H., Audic, S. and Claverie, J.-M.

Selfish DNA and the origin of genes

JOURNAL Science 291 (5502), 252-253 (2001)

REFERENCE

3 (bases 1 to 14117)

Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.

Mechanisms of evolution in Rickettsia conorii and R. prowazekii

JOURNAL Science 293 (5537), 2093-2098 (2001)

MECLINE 21442074

PUBMED 11557893

REFERENCE

4 (bases 1 to 14117)

Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.

Direct Submission

SUBMITTED (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France

A public version of R. conorii genome database is accessible at <http://dis-server.cnr-mrs.fr/>. The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.

location/Qualifiers

1..14117

/organism="Rickettsia conorii"

/strain="Malish 7"

/db_xref="taxon:781"

127..1809

/gene="y1dC"

/note="RC0074"

127..1809

/gene="y1dC"

/codon_start=1

FEATURES

source

gene

CDS


```

/transl_table=1
/product="60 kD inner-membrane protein"
/protein_id="AAL02612.1"
/db_xref="GI:15619112"
/transl_table="MNNNTLNLIAILLLSLIFGWOYFVVKPEQKKOQOIAVQKAE
NLKQQLKALVEPATGIYQESQVQRIKIESESLTGSISLKGRLPDDLILKKQDL
SKNPEVRLFSNPANTENAYFAEVLGNLSLKVLPNDITWNSDSEILSEKPYHLFW
VNEDEVFLVITVDENYLFTEQITVNSDKELPYOYGLINRKYIAVEKAVILHQ
GPTICIDENKEYSDIDIKDKSEKFAKSVDMGTIDKYWLSLIPDKSSNSNPN
YALKQCTERYQVDPISPVQIIRKGENSISKRIIPGAKKYVDLDEKTEKQIDKLPDA
IDEGMTYITRKPEYAMNFFYGVNGVSLITVYTIKILMFLYLANKSYRSMKMN
LQPEIDRIKYLISDDKARLNQEIYALYKREKYNVACCLPLVQIPVFFSYKVLVT
IEMRQAPFYGIWKIDLSASDPTIENFLGLLPSPSPCLMIGAMPILMAITMFLQOKKS
PEPADPMQAOYMKEMPLIFLFMFSSFPVGLLIYMSNNIISIIQOYIINKFN"
1813..2358
/gene="pgsa"
/ncore="RC0075"
1813..2358
/gene="pgsa"
/codon_start=1
/transl_table=1
/product="CDP-diacylglycerol--glycerol-3-phosphate
3-phosphatidylintransferase [EC:2.7.8.5]"
/protein_id="AAL02613.1"
/db_xref="GI:15619113"
/transl_table="MRIDKNLPNYLTARIMVPIVITLVFVYINNSLARTIGALLFVLA
SIMPFGYIARKKNYLVSRCGMEDPIADKILVGCYIIMLKKNDVDEIPCLILARE
FLVSGLEPFLALVSVPSRLAKVTFLOWFALSIILLSKSGSGLIYLDIVGETIIM
IAFLTITGTSYKACKRYF"
2372..2932
/gene="RC0076"
2372..2932
/gene="RC0076"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL02614.1"
/db_xref="GI:15619114"
/transl_table="MIHVYIIGFIWSSNLVYAANNDIANKYKVEVSEHKT
LNEASTLISCVDFRLIDETDKLKKQJGLEENFKVSLPGASLILNDKTIYWKTE
DTIILLOELHNIKOIYFLDRECGAVYILIGEOELNKEKETAAHAAILNKARDIIE
KIPQLKYVTEFLGDGVVEOYEIIPS"
complement(2964..3063)
/ncore="REP01, repeated element"
complement(3102..3521)
/gene="RC0077"
complement(3102..3521)
/gene="RC0077"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL02615.1"
/db_xref="GI:15619115"
/transl_table="MVLRSRLRKYMIKKIIFGIALILSTCFANSTYSDGSKDAA
KTNDIITOKIIDDESAYAGTKEPEVREIOKYRAIVAIINKKRELYNRLSKAQNFL
AEQOKYKOKLSISKLIVENDOKNNTADSNKSKDTR"
complement(3542..3685)
/ncore="REP05, Rickettsia palindromic element 1"
3705..3977
/gene="RC0078"
3705..3977
/gene="RC0078"
/codon_start=1
/transl_table=1
/product="similarly to Birc"
/protein_id="AAL02616.1"
/db_xref="GI:15619116"
/transl_table="MYFLFVCYCLVSIINILVGNIOYNNNAQEFYSRTINADLSN
YKATSYLPBQAHIVDAGCGVCDTRKFLSORVYATFADSTENK"
4544..5392
/gene="RC0079"
4544..5392
/gene="RC0079"

```

```

/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL02617.1"
/db_xref="GI:15619117"
/transl_table="MAYRTNANAIVNGFPEIAGSDGRLTLMIDSKLFSRLKQJ
SLIIDIQNIQITKASAKILVIDKSVINQVYFNLNIDIFYNNVMASTLTSYT
NKELLIDQNSIVTASAKILVSKHGNOIPQKGFILSPQATSLPVNINDSVKLILEFTD
KDGKLINLSTASTVYTIPLLVQGNKVVNDPKDDAHARTALGVCNDGTIYIYVE
HITKQHVADKILVOVRSILRQEKINVDKLIIPALKILERHLVNTVIGLTYTELAD
YMLTWI"
complement(5609..6283)
/gene="def3"
/ncore="RC0080"
complement(5609..6283)
/gene="def3"
/codon_start=1
/transl_table=1
/product="polypeptide deformylase [EC:3.5.1.31]"
/protein_id="AAL02618.1"
/db_xref="GI:15619118"
/transl_table="MSRLKTAALLTAIIFITGCKSNMKNENNISLPQYVTLNIQTLS
SEDKTIRIKAKTLNPLSEDLRDISILKRYDQEEKACGLAAPQIISCIITFAVH
EADLKKMHPDLKDTMPTKTWINPSYKPIGIDKHEDYEGCFSVNATGPVAREKKIH
HAYDINGNQJGIAEGFLAVYIQHEIDHLNGKVLQDVAPKKIKMTKEYLEMKKAME
QENIKS"
6482..7978
/gene="tlc1"
/ncore="RC0081"
6482..7978
/gene="tlc1"
/codon_start=1
/transl_table=1
/product="Adp,ATP carrier protein"
/protein_id="AAL02619.1"
/db_xref="GI:15619119"
/transl_table="MNNPKNDNYLSELSKYIMPIERYENKFLPMAPMPCILLNST
LRSIKGFEVVDIGAEALISPLKTYIVYPSAVIAMVYIKYICIDKQNVVYETSPEL
GYALFAPVLYPRDYLVHPDETIESNSVAVPNVKPIIRYGGKRSFSTYTMALSKT
MMLSLTFWQFRANQITKIDEAKRFYSMGLANLALPVTSVITIGCLHEKQIYAEHLK
FVPLVYALMITSFVLITFRMMKNVNLIDRLDPALVKKRAKAKMSLIDSEFMIFT
SKYVGYTALLIAYGVSVNVEGVKSVRELYETKDAYTIYMGKFOYIGVVAIAM
LIGSINILRKYSWILTAAMITPLMAVITGAEPFAIFEPDSVIAHLTGILASGPLAVM
IGMIONLNSGVKYSFLDPTKMAVITPLDDKIDRYKGAAYEVJGGRGKSGALIOST
FFILFPAEPYEAIPYASIFEVYIIMITYAVKLNKEYVLVAKTEK"
8175..9470
/gene="g1pt"
/ncore="RC0082"
8175..9470
/gene="g1pt"
/codon_start=1
/transl_table=1
/product="glycerol-3-phosphate transporter"
/protein_id="AAL02620.1"
/db_xref="GI:15619120"
/transl_table="MWTNRKIKTPRSRPYKKVYKNSMVRILYSIIIGYATPYPCRONF
NIATPAIRETFGVTKIQIGWILTASSIMYGVKACNGFISDKVNARLFWVLGLFVGI
ITMLIGFSDSLMILGILMIASNMFQSGMWEPATKMLHWFASKELGTRKMGATSNQI
GGALAMISCGYLIDKPFMRAAFVYGVVACVSLFNLHRLNSPKRSEGLSTVEYKSY
PLEAIGDYERKLIPQLKAVCNKLIYVCANNEYIIRSGVYMAPTPLKDLRNS
LANGLQIGLYEMTIGIPGALIAVSLRKLQGRGVPASVTCMVLSILILYFMPKLPFH
SELSTYVILSLIGFFVSGPOLYAGVIAAADSTRAVGTANGLSGFLGALAAAGVY
GWISDNYGMNGVLFETISALLGGFLALIMNRSKRYT"
9495..9917
/gene="ndk"
/ncore="RC0083"
9495..9917
/gene="ndk"
/codon_start=1
/transl_table=1
/product="nucleoside diphosphate kinase [EC:2.7.4.6]"
/protein_id="AAL02621.1"
/db_xref="GI:15619121"

```



```

/protein_id="CAA64329.1"
/db_xref="GI:1707364"
/db_xref="SPTREMBL:P92935"
/translation="MEGLIOTRGLISLPASHRSKVLDPSSHGLKORLETTNLPALISLS
LMTNRNRPESKSHLGRFPTRREHEDSLARKLRPRKRCVDEEDTAMAVSPRTIG
VEVTLTKRIVPLGLMFCILFNVTILRKTQVLLVTAAGSAEILPELTVNPMAL
GFMLLYPLKLSNLSKALFVIVPFIYFPAFGVMPRNLQIPELALAKLLATLG
PRFMPLAIRMVSWKFLFYVMAELMGSVYVLFPMFQVNOITTYDEAKFPLFGLGA
NVALIFSGRTKVESSNMNRNLGPGVDGVAIVKAMSVYMGILACIEMPMVNRVYP
LPBRSKKKVKPQMKMESLKFVLSPIRLATLIVAYGISINLVEVTKMSKLKSP
PSPDEYSFMDGDFSTCTGATFATFMLLSQYVEKRYGQWVAKITPVTLLGVAFSL
ILFGFPAIPLAKLGMPLLAAYVYVPEVSARVOHSSSTPSMODCLPDEVSK
VKAKIOLMMSATIKSGGALLQOFILFISLANSTPYLGLILGIYVAMLAASKLE
GPV"

```

BASE COUNT 540 a 456 c 503 g 640 t
ORIGIN

Query Match

17.3% Score 283.8; DB 8; Length 2139;

Best Local Similarity 52.8%; Pred. No. 3.3e-57;

Matches 749; Conservative 0; Mismatches 637; Indels 33; Gaps 5;

```

QY 120 CTAAGAAAGTCTGCCAATGTTCTCTGATGATTAACATTAACATACGGTG 179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 403 CTGAAGAAAGTGTCCCTTAAGGCTTAATGTTCTTTCATCTTTCATTAACAATC 462
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 180 TTACGGATACAAAGACACTTAATGCGAGGCTCTGTTCTGTCGACAGAGCAATA 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 463 CTTAGGACACGACAGAGATTTTGGTGTGACGGCTAAAGAAAGTTCTGCTAGATTA 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 240 CCTTCATCAAGTTTGGCTTGTGCCCTGCTGATTAATCTTAATGCTTAATATGCA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 523 CCTTTTGAAGACATGTTGGAATGTTCCGATGCTATGAGGTTATGTTGCTATACACC 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 AAGCTAAGTAATATTTAAGTAAGACAGGCTTATTTATGACGTGGAAAGCCCTTTTA 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 583 AACTTTCCAAATGTTCTCCAAAAGGCTCTTTTACAGCTTAATGTCCTCTTCATTT 642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 ATTTCTTTGCCCTTCCGACACTTAATTTATCCGCTACGGAGTTTATACATCTACA 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 643 GCTACTTTTGGAGCCTTTGGTTCGTGATGACCCCTCGACAGAAATTTATTAAGCTGAA 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 420 GAATTTGCTGACCGTTTACAGGCATCTACCTCCAGATTTGCTAGACTCGTTGCATC 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 703 GCTCTTGTGATTAAGTTCTTGCACACACTCGGCCCAAGTTGATGGGCTCTCGCAATC 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 480 TTAAGAACTGACATTTGCTCATTTTATGACTTGTGAACTATGGGAAGCTCATG 539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 763 ATGAGGATTTGAGTTTCTGTTGTTCTATGTCATGAGCTTTGGGCTGATGTTGTC 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 540 CTATCTTAATGTTGCGGATTTGCTATGAATTAACAAAATCCAGACCAAGGCT 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 823 GTTTAGTTCTTCTTGGGGATTTGGCCACGACGATTTACACAGTTGACAGACCAAAAG 882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 600 TTCTACGCTTTTGGGATTCGAGACTAATATTTCTTAAGCTTCTGCTGCTGCAAT 659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 883 TTCTATCCCTGTTTGGATTTGGGCAATGTTGCACATATCTCTCAGAGAAGACTG 942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 660 GTTTGGGCTTCAAGTTGAGACTTCCCTTCTGAAGTGTGATGCTTGGGGAATTTCT 719
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 943 AAATATTTCTTAATGAGAAAGAAATCTTGCTCGAGGATGAGGCTGTTTCA 1002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 720 TTACGCTTTTGGGATTCGACTAATTTGATGCTGACTTCTTATGCCAGTTACTG 779
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1003 TTAAGAGCATATGATGATTTGCGGGGATGCGTCCGACATCTGTTTCTCTACGTG 1062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 780 TGGATCAATTAAGACGATTTAGCCGATCTCCCTCTATATTCAGAGAAGAAATGCAAAAG 839
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1063 TGGGGAATAA-----GATATGTGCCCTTCTCAACCCGTTAGC 1098
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 840 GGGAAAAAGGTGTAACCTAAATGATATGAAGAGATGCTTCTCTATCTGTATGA 899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1099 AAGAAGAAAGGTGAACCAAGATGGAACAATGAGAGCTTGAAGTTCTTGGTGCA 1158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 900 TCTCCTTAATATCTTTTATTAATACCTCTGATTTATGCTATGATTTGACATTAACCTTA 959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1159 TCACATATCATTAAGGATCTTCTACTTTGGTGGTGCATATGAATTAAGTATCAACCTT 1218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 960 ATCGAAGTGACTTTGAAAAAGTCAGTGAAACATGCAATATCTTAATATGATGACTATAGT 1019
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1219 GTTGAAGTCACATGGAATAATCAAAAGCTTAAAGTCAGTTCCCTAGCCCAAGAAATATCA 1278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1020 GAGTTCATCGGGAAGACTTCTCCTTCTGAGACTGCGGATAGTATCCGTAATATCATGCTATT 1079
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1279 GCAATTAATGAGGACCTTCTCAACCTGACACAGGATATTGCAAC--ATTCACAAATGATGCT 1335
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1080 GTTGTGGAACGATATTCGTAATATTTGATGATTAACGTGAGCCCTAGTACACTCTGTC 1139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1336 CTAAAGCAATACGTTGTTAAGAAATATGTTGGGAGATAGCTGCAAAAGATCACACCAAC 1395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1140 ATGATTCCTTAACAGATATGCTTTCTGCTCTGTTATCTTTAGAAAACCAAGCTCT 1199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1396 GTTCTGCTATTAACCGGTGTCGCTCTTCTCTGATACGTGTTGGGCGCCATTCGCA 1455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1200 GGGCTGCTGCTATGTTGCGGTACAACTCCTCTCATGCTAGCTGTGTGTGAGACTATA 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1456 CCATTGTTGCCAACCTTGGTATGACACCGCTACCTGACAGCTGATGACTTGTCCCTCA 1515
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 CAGAAATATCTTTGCAAAATCCACAAATATACGCTCTCTTTGACTC--AACCTAAAGAAATG 1316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1516 GA-AGTATCTTACGACAAAGTCCAAAGTACAGCATTTCTTGACTCCCTCGCAATGACAGA 1574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1317 GCCTATATCCCTTGTACCAAGAGCAAAAAGTCAAAAGGTAAGCTGCTATTAATGATAGT 1376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1575 ATGCTATATTCATTGATGATGAGGTCTCCAAAGGTAAGGCAAGCTGCAATATGATGCT 1633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1377 GCGCGCCGCTTGGAAAAATCAGAGAGAGCTTAAATCCAAAGATTTGCTCTATCTGT 1436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1634 -CTGAACCATTTGGGAAATCAGAGCGGTGCTTAATCCAGCATGATGATCTCTTACATTC 1692
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1437 GGAAGTATTTGAGACTATGACCCCTATCTTGCAGATATCTTTCATCAATTCGATTT 1496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1693 GGGTCACTGCCAATTTCCACACTTACCTTGAAGTACATCTCTGTGGTATATGACTGCA 1752
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1497 TGGTGTGTTTTCGCACTAAGTTAAACAACATTTCTTA 1535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1753 TCGTTAGCAGCAGCTAATATGCTGGAGGACCAAGTTTAA 1791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 26
AF087957 1505 bp DNA Linear SYN 07-NOV-1998
LOCUS Synthetic construct ATP/ADP translocase (tlc) gene, complete cds.
DEFINITION AF087957
ACCESSION AF087957
VERSION AF087957.1 GI:3834578
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequences.
1 (bases 1 to 1505)
REFERENCE
AUTHORS Alexeyev M.F. and Winkler H.H.
TITLE Gene synthesis, bacterial expression and purification of Rickettsia prowazekii ATP/ADP translocase
JOURNAL Unpublished
2 (bases 1 to 1505)
REFERENCE
AUTHORS Alexeyev M.F. and Winkler H.H.
TITLE Membrane topology of the Rickettsia prowazekii ATP/ADP translocase
JOURNAL revealed by novel dual pho-lac reporters
3 (bases 1 to 1505)
REFERENCE
AUTHORS Alexeyev M.F.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1998) Microbiology and Immunology, University of South Alabama, LMB Building, Mobile, AL 36688-0001, USA
FEATURES
source
1..1505
/organism="synthetic construct"

/db_xref="taxon:32630"
/note="coding region for Rickettsia prowazekii ATP/ADP translocase engineered to optimize expression in E. coli and to introduce unique restriction endonuclease sites in every region coding for a potential hydrophilic loop (except loop 11) thus creating a cassette version of the gene"
4..1500
/gene="tlc"
4..1500
/gene="tlc"
/function="transport"
/note="obligate exchange system"
/codon_start=1
/transl_table=1
/product="ATP/ADP translocase"
/protein_id="AAC72100.1"
/db_xref="GI:3834579"
/translation="MSTSKSENYLSELRKIIMPEYENKKEFLPLAFMNCILNYS
LRISIKDFVVIDIGTESISELKYIVLPSAVIAMIVKLCIDILKQENVEVITSEFL
GYFLAFVLYPYDVLVPHDKITSESLAYPNKWEIKITGKWSFASYIADLMGT
MLSLFMQFANOITKIAEAKRFYMEGLANLALPYTSVIGVLEHKOIVAEHK
FVPLFVIMITSSFLIILTYRMNKNVLDPLVDPALVKEKTKAKISPLESKMFT
SKYGYIALILIAAGSVNLVEGYMKSKELYPTKAYITIMQEPFOGVIAIAM
LIGSNILRKVSWLTAAMITPLMFTGAAFSEFSDSVIAMNLGLIASSPLVLA
IGMLQNLVSKGVKISLEDAIKMNAVILPLDRLRVKGAAGAAVEVIGRLGSGGALIOST
FFILFPEVGEAFPEAFSIFFIIVILMIFRAVKLNGEYOLVKNKEK"

BASE COUNT 346 a 389 c 297 g 473 t
ORIGIN

Query Match 16.8% Score 274.2; DB 12; Length 1505;
Best Local Similarity 52.0%; Pred. No. 6.7e-55;
Matches 783; Conservative 0; Mismatches 688; Indels 36; Gaps 6;

66 GAAAACCTTTTGGAAATGCGCTCTTCTTGTGGCCGATACATCAGAGCTAAG 125
11111 11111 111111111 11111 11111 11
22 GAAAACCTACCTGTCTGAACCTGCGTAATCATCTGCGCATGCAACATACGAAACAA 81
11111 11111 111111111 11111 11111 11
126 AAGTTCTGCAATGTTCTTAATGTTCTGTATTAATTAACATACGATGCTTACGC 185
111111111 111111111 111111111 11111 11111 11
82 AATATCTGCGCGGTTCATGATGTTCTGCAATCTGCTGAACCTACTCATCTGCGCT 141
111111111 11111 111111111 11111 11111 11
186 GATACAAAAGACATCTATTGTGGAGCTCTGCTGTGTCGACGAGCAATACCTTTC 245
111111111 11111 111111111 11111 11111 11
142 TCTATCAAGACGCTTGTGTTGT-----TACCGATACGATCGAATCATCTCTTC 195
111111111 11111 111111111 11111 11111 11
246 ATCAAGTTTGGCTTGTGTCCTGCTGCTAATTAATCTTAATGCTTAATTAAGCA 305
11111 11111 111111111 111111111 11111 11111 11
196 CTGAAACCTACATCTGCTGCGCTGCTGCTATGCTATGATCATCTACGTTAACTG 255
11111 11111 111111111 111111111 11111 11111 11
306 AGTAAATATTTAAGTAGAGAGCCCTTAATTAATGAGTGGAGAGCCCTTTTAATTT 365
11111 11111 111111111 11111 11111 11111 11
256 TGCGACATCTGAAAGAGAAAGATATTAACGTTATCACTCTTCTTCTCTGCTGCTAC 315
11111 11111 111111111 11111 11111 11111 11
366 TTTGCGCTTCCGACATGATTTATTCGCTACCGCATGTTTATCATCTACAGAAATTT 425
111111111 11111 111111111 11111 11111 11111 11
316 TTGCGCTGCTGCTGCTTCTGCTGCTACCGCATGCTGCTGCTGCTGCTGCTGCTGCT 375
111111111 11111 111111111 11111 11111 11111 11
426 GCTGACGCTTACAGAGCATCTACCTCCAGATTTGCTAGAGCTGCTGCTGCTTAA 485
11111 11111 111111111 111111111 11111 11111 11
376 ACTATCGAAACCTGTCTGCTGCTTACCGCAATTTCAAAATGTTTCAATCAATGCT 435
11111 11111 111111111 11111 11111 11111 11
486 AACTGACATTTGCTGCACTTTTATGTAATCTGCTGCAATGAGGAGGCTGCTATCT 545
11111 11111 111111111 111111111 11111 11111 11
436 AATGCTGCTTCTGCTTCTTCTTACATCTGCTGCAATGCTGCTGCTGCTGCTGCT 495
11111 11111 111111111 111111111 11111 11111 11
546 CTAATGTTTGGGATTTGCTAATGAATTAACAAATTCACAGAGCAAGCGTTTCTAC 605
111111111 11111 111111111 111111111 11111 11111 11
496 CTGCTGTTTGGGATTTGCTGCTAAGATCACTAAATCGCTGAAGCAAGCGTTTCTAC 555
111111111 11111 111111111 111111111 11111 11111 11
606 GCTCTTTCGCTATCGAGCTAATATTTCTTACAGCTTGTGGCGCAATGTTTGG 665
111111111 11111 111111111 111111111 11111 11111 11
556 TCTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 615
111111111 11111 111111111 111111111 11111 11111 11

666 GCTTCAAGTTGAGAGCTTCCGTTTCTGAAGGTGATGATCCCTGGGGAATTTCTTACGT 725
11111 11111 111111111 11111 11111 11111 11
616 TTTCCTGACAGAAAACCT-----CAGATCGTTGCTGAACACCTTAATGTTGCT 663
11111 11111 111111111 11111 11111 11111 11
726 CTTTGTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 785
11111 11111 111111111 11111 11111 11111 11
664 CCGCTGTTGCTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 723
11111 11111 111111111 11111 11111 11111 11
786 AATAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 845
11111 11111 111111111 11111 11111 11111 11
724 AACAAACGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 783
11111 11111 111111111 11111 11111 11111 11
846 AAGGCTGCTAAACCTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 905
11111 11111 111111111 11111 11111 11111 11
784 A-----CTAAGCTAAACCTGCTTTCATGCAATCTGCAAAATGATGATGATGATG 837
11111 11111 111111111 11111 11111 11111 11
906 TATATCTTCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 965
11111 11111 111111111 11111 11111 11111 11
838 TACGTTGCTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
11111 11111 111111111 11111 11111 11111 11
966 GTGACTTGGAAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1025
11111 11111 111111111 11111 11111 11111 11
898 GGTGTTTGAATATCTAAGTTAAAGATGATGATGATGATGATGATGATGATGATGAT 957
11111 11111 111111111 11111 11111 11111 11
1026 ATGGGAACTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1085
11111 11111 111111111 11111 11111 11111 11
958 ATGGGTGCTTCCAGATCTTACAGAGGTGGGTGCTGCTGCTGCTGCTGCTGCTGCT 1014
11111 11111 111111111 11111 11111 11111 11
1086 GCTAAGCTGATCTGATTAATTTGATGATGATGATGATGATGATGATGATGATGATG 1145
11111 11111 111111111 11111 11111 11111 11
1015 TCTAATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1074
11111 11111 111111111 11111 11111 11111 11
1146 CTCCATACAGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1202
11111 11111 111111111 11111 11111 11111 11
1075 TTTCATACAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1134
11111 11111 111111111 11111 11111 11111 11
1203 CTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1262
11111 11111 111111111 11111 11111 11111 11
1135 CTGACCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1194
11111 11111 111111111 11111 11111 11111 11
1263 AATATCTTCTGAAATCACAATAATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1322
11111 11111 111111111 11111 11111 11111 11
1195 AAGCTTCTGCTTAAAGGTAAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1254
11111 11111 111111111 11111 11111 11111 11
1323 ATCCCTCTTGAACCAAGCAAAAGTCAAAAGGTGCTGCTGCTGCTGCTGCTGCTG 1382
11111 11111 111111111 11111 11111 11111 11
1255 ATCCCGCTGCAACAAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1314
11111 11111 111111111 11111 11111 11111 11
1383 CGCTTGGGAAATCAGAGAGGCTTTAATCCAAAGGTTGCTGCTGCTGCTGCTGCTG 1442
11111 11111 111111111 11111 11111 11111 11
1315 CGTCTGGGTAAATCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1374
11111 11111 111111111 11111 11111 11111 11
1443 ATTGAGAGCTAT-----GACCCCTATCTTCTGCTGCTGCTGCTGCTGCTGCT 1456
11111 11111 111111111 11111 11111 11111 11
1375 TTGCGTTTCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1434
11111 11111 111111111 11111 11111 11111 11
1497 TGTGTTGTTTCTGCAACATTAATTAACAAATTAATTAATTAATTAATTAATTAAT 1556
11111 11111 111111111 11111 11111 11111 11
1435 TGGATCTTCTGCTGTTAAAGCTGCTGCAACAAAGATTAACAGGTTCTGCTTAA 1484
11111 11111 111111111 11111 11111 11111 11
1557 CAAGAG 1563
11111 11111 111111111 11111 11111 11111 11
1495 AATGAG 1501
11111 11111 111111111 11111 11111 11111 11

RESULT 27
AY120885 1900 bp DNA linear BCT 31-JUL-2002
LOCUS AY120885
DEFINITION Holospora obtusa non-mitochondrial nucleotide transport protein
(nnt) gene, complete cds.
ACCESSION AY120885
VERSION AY120885.1 GI:22035402

RESULT 28
AE001323/C
LOCUS
DEFINITION Chlamydia trachomatis section 50 of 87 of the complete genome.
ACCESSION AE001323 AE001273
VERSION AE001323.1 GI:3328931
KEYWORDS
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis;
REFERENCE Bacteria; Chlamydiales; Chlamydiaceae: Chlamydia.
AUTHORS 1 (bases 1 to 24297)
Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
JOURNAL Science 282 (5389), 754-759 (1998)
MEDLINE 99000809
PUBMED 9784136
REFERENCE 2 (bases 1 to 24297)
Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
Comparative genomes of Chlamydia pneumoniae and C. trachomatis
Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 3 (bases 1 to 24297)
Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Direct Submission
Submitted (20-May-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
Source Location/Qualifiers
1..24297
/organism="Chlamydia trachomatis"
/strain="D/UW-3/CX"
/db_xref="taxon:813"
complement(83..1705)
/gene="CT495"
complement(83..1705)
/gene="CT495"
/codon_start=1
/transl_table=1
/product="ADP/ATP Translocase"
/protein_id="AAC68096.1"
/db_xref="GI:3328932"
/translation="MSSEVKSFSKFRGYFPPIYVSERSKPIPLFPLAEFGVNVALLK
TTKDSLVLVGRAGAEEVPLKRWGIVGAVIVMIGWMSRRYSRGIVFTSLVGGFL
GEFALFATVYIPIDALHLNKLAKLOSILPPGSRGFVWQVWSYSLYVMSSELMSS
IVSLTEFGVANHITSVREAGRFALINTGNTSSVFEAGVSLMGRNPVLAFFMAVD
PMHEVLNITLITLIVAGVILYLYOKDLRMDETSMLEGLAEMSVQLEKRRSK
AKASLFAILLRSRYLIGIAVIVSYNVIHLFEVVKDQCYARASVENSNMSRT
TTLTGIVSALTGIFAAAGOTIRKMGTTGALVPLTMTLTGALFPGATYAVAGDMITG
GITLISPLVITLWAGVQVNFSAIKFPEQDTENAEIPLDEDEKNKGAIDGVIS
RVNSEDVLOEEREASLVDAESREEBVTLL"
complement(1917..2423)
/gene="pgsa_1"
/note="CT496"
complement(1917..2423)
/gene="pgsa_1"
/codon_start=1
/transl_table=1
/product="CDP-diacylglycerol--glycerol-3-phosphate
3-phosphatidyltransferase"
/protein_id="AAC68097.1"
/db_xref="GI:3328933"
/translation="MROCNLTLSSRWALLFCOERTITRLVIFAAWSDVLDGYL
ARRYNATSRIGSIDPADDKITFELICGVLEWMSIGLTHALIFSRDILVFEGFL
SWRGMKGYDYRALSFCKFTVVOFLIFGNTIGMEKPEVLMALPLVILGALYLERVL

1
gene complement(2610..2693)
/gene="rRNA_Ser_2"
complement(2610..2693)
trna /gene="rRNA_Ser_2"
/product="rRNA-Ser"
gene complement(2819..2968)
/gene="CT496.1"
complement(2819..2968)
CDS /gene="CT496.1"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAC68130.1"
/db_xref="GI:3522901"
/translation="MGNGFFVVAIVRCLFLMLVKALASRLGIRGITPOBERGEEK
YRPLG"
2937..4355
gene /gene="anab"
/note="CT497"
2937..4355
CDS /gene="anab"
/codon_start=1
/transl_table=1
/product="Replicative DNA Helicase"
/protein_id="AAC68098.1"
/db_xref="GI:3328934"
/translation="MANTKKKPPDPLSLPNSKSEMTVIGMLTSVNHNLNANLL
QEDPPYFLEHRIIRVLODAKSDRPNDPHLTGEBLKRDDLNTIGGFSYITLSEFA
GTSAYIEYATIRKSLTRKMDQAKTEKKAEEPRDVTALDDQNLFLRISOTT
MLAPHLVADKLKAVASSKDKSFLIALQERKAEQASAHSDSSMLSGFPHFLDLK
MISGFSPSNLITILARPAMGKTALANIVENCFDSAPVGIFFLEMTVDDLIRIIC
SRSEFEAKRISVDISGRDFORVSVVMEHEHTLIDYVGLKITDRLARARAKMS
YDIOFLVIDYLOLSSSGNLNSDSRNOEISEIRMLKNLAREINIPILCSLSRY
EDRANHPRIMSDLRSSGISDQDQIMFLAREYDPRDKGTALLIYAKNRHSIGS
VOLYERKDFARFRNYACCEPFG"
4649..6481
gene /gene="gida"
/note="CT498"
4649..6481
CDS /gene="gida"
/codon_start=1
/transl_table=1
/product="FAD-dependent oxidoreductase"
/protein_id="AAC68099.1"
/db_xref="GI:3328935"
/translation="MHTFPVDYDVIVIGAGHAGCEAAVCAAKMGASVILLTSLNLDIA
KLSCNPAGVIGKGIHYREIDALGGINAETDLSGIOFRILNOKGRAVAPRAQVYK
QLVITHMKRLLEQVPGHLIWGTAEALDNGEVLGVSTKEGNAVLGKTVVLSGTFM
RGLIHIGTQNFSGRGLDAAASLGISEDLRKRGFLGRKTGPARLASSIDFVME
QPGDHNCVFRNEMFVPTLPVSGCHLTHTDQTKDLTKMLHRSALVGGRIEAGVR
YPSIIEKIVFADKDRNHIEPEEGNLTCVVYNGISTSMFPVOYDITRSVGLEN
AIIIRPAVAYEDVHGVNIPPSLESKLIEGLFCGGNGTGYEEAAAOGLIGVNA
VXKYLRIHPPEVROESYIGWMLDITTOVDEPERMFTSAEHRLLTRDNACMRS
HNGSLGLSSERAPMOEQKACIEQEKERSKTFKRYGDIYVPLTYLCEPEVSYO
LTFEPADVRDLPVAGASLEMEKITSQYISRQDTLRMSRSENISIPEDIDVHSIS
ALSLAEARKSKFTPRTIGSAAISGISVADIOVLWVSLKKDAH"
6471..7172
gene /gene="lpla_2"
/note="CT499"
6471..7172
CDS /gene="lpla_2"
/codon_start=1
/transl_table=1
/product="lipoate protein ligase"
/protein_id="AAC68100.1"
/db_xref="GI:3328936"
/translation="MLNCVFNHCEGPIRKQQLQEBALRTSSQNCFLVNLPEAV
VLGISRKPERDLHVEHLKEDGIPITIRRYSGGGVFLDADSLMWSMIINSPTSPSSKD
LQWTDODIYAPIPTGFKITENDYIDPDKIKGNAOYIOKRYMWHTTFLMNNPKKL
ARYLPPEIOPSVYNSHDEFLTITIELDSDRDFLSQLKQSAASKMWEGSGIQL
TPMISLHRRATQIL"
complement(7229..7654)
gene

```

CDS
    /gene="ndk"
    /note="CT500"
    complement(7229..7654)
    /gene="ndk"
    /codon_start=1
    /transl_table=11
    /product="Nucleoside-2-P Kinase"
    /protein_id="AAC68101.1"
    /db_xref="GI:3328937"
    /translation="MEORTLSTIKPDSVSKAHGELIAIFESKGLIAAMKXVHLSYKE
    AEGRTVYHKERPFQELVDENISGPVVMVLOGEAVAPRNRLMGATNPKKEAESIR
    ALPESIGVNAVHSDSLNNAIEVSYFFAKTEIVNSVA"
    complement(7814..8416)
    /gene="ruva"
    /note="CT501"
    complement(7814..8416)
    /gene="ruva"
    /codon_start=1
    /transl_table=11
    /product="Holliday Junction Helicase"
    /protein_id="AAC68102.1"
    /db_xref="GI:3328938"
    /translation="MYEIKGTLTHTINESYVIESFGIYATIMSERPLVDLRPMHQ
    EVLIYVSHIRETEHVLYGFSRARECEFRLLISGIGPKGLSLNMFPLQELCSI
    ARLENVKAIASVPGIGKKTAEKLMVDLQKLPTELLLEPVVSVSTANSSEFKEGIG
    ALMNLGFRLAADRMTEAVKEELSEASVALLPALKRS"
    complement(8436..8948)
    /gene="ruvc"
    /note="CT502"
    complement(8436..8948)
    /gene="ruvc"
    /codon_start=1
    /transl_table=11
    /product="Crossover Junction Endonuclease"
    /protein_id="AAC68103.1"
    /db_xref="GI:3328939"
    /translation="MADLMGIDPSTLYCGVALLIKENRYHIIHPHSGKVKLSOKLAT
    AHRYKQLEIESTILOESPRAVVLQYVIRPQSTIKLGARGVLLIASLOQV
    FEYAPNTAKKAAGVGNASKKQVOVLMVSKLRVPLLAEDNEDIDATRALAMCHHLA
    PYQDKLTLY"
    complement(9056..9610)
    /gene="CT503"
    complement(9056..9610)
    /gene="CT503"
    complement(9056..9610)
    /gene="CT503"
    /codon_start=1
    /transl_table=11
    /product="Hypothetical protein"
    /protein_id="AAC68104.1"
    /db_xref="GI:3328940"
    /translation="MWMLSVLRLRLSLRFRKVRAPTSPOYEYSRILOFLMLK
    DSDIYEMTEOMLEFGASKSEDLIRKKKMOIOSAKEOSILAMBOOVCEIKSOLA
    SOENANOEISKLOAEHNHLOLRLEAKLQOAHNDVIDELRDLVESVOOMVSEGR
    RLCTEHKIRLVEQIDRLAKEV"
    complement(9695..9777)
    /gene="tRNA-Leu-3"
    complement(9695..9777)
    /gene="tRNA-Leu-3"
    /product="tRNA-Leu"

Query Match      15.7% Score 256.4 DB 1: Length 24297;
Best Local Similarity 49.7% Pred. No. 9,3e-51;
Matches 749; Conservative 0; Mismatches 736; Indels 21; Gaps 3;

```

```

Db 1578 TAAGACCAAGAGAGACCTCCCTGTTTATGTCGGCTAGAGCAGGCGGAAGCATAC 1519
QY 241 CTTTCATCAAGTTTGGCTTTGTGCCCTGTGCTATATCTTATGCTTATTTATGCA 300
Db 1518 CTTTCATCAAGTTTGGCTTTGTGCCCTGTGCTATATCTTATGCTTATTTATGCA 1459
QY 301 AGCTATATATATTTAAGAACAGCAGCCCTTATTTATGACAGGGAAGCCCTTTTAA 360
Db 1458 GATGAGCTGACGCTTATTCAGAGGACCGCTGTTATCTTATGTCGGCGGTTTTTAA 1399
QY 361 TTTTCTTTGCCCTGTTCCGAGTGAATTTATCCGTCAGCCGATGTTTACATCCACAG 420
Db 1398 GCTTTTTCGCTGTTCCGACAGATGATTTATCCATAGAGATGCGCGACATCTAATA 1339
QY 421 AATTTGCTGACCGCTTACAGGCCATCTCTACATCCAGAGATTGCTAGGACTCGTCATCT 480
Db 1338 AATTTGCGGCAAAACATCAAGTCATCTCTACCTCGGGGGGAAGAGGTTTGTATGTCATG 1279
QY 481 TAAGAACTGAGCATTTGCTGCAATTTATGACTTCTGTGAACATATGGGGAAGGTCATGC 540
Db 1278 TTCAATACAGGAGCTATAGCCTGATATATGTATGTCTGATGATTTAGAGTTTCCATAGTTT 1219
QY 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATCAAGAACGATTT 600
Db 1218 TGTCTACTGCTGTTTGGGGGTAAGCCACATATTACGAGTGTTCGGAAGCAGGGCGGT 1159
QY 601 TCTACGCTCTTTTGGTATCGAGCTAATATTTCTTACTAGCTTGTGCTGCATTTG 660
Db 1158 TTTACGCTCTTATTAATACGTGATTAACCTCTCTCTGTTTTCGCGAGAGGTTTCTT 1099
QY 661 TTTGG---GCTCAAGTTGAGAGCTTCGTTTCAAGAGTGAAGTCCCTGG----- 710
Db 1098 TGTGCTCGGTAAATTCCTGTAATTCCTTCCCTATGCGCGTAGATCCTTGCGATGAGG 1039
QY 711 -----GGAATTTCTTACGCTTTTATGATGAGCTATGACTATTTGATCTGCACTTGTTC 762
Db 1038 TGTGCTCAATACATCACTGATCTATGCTGTGCGCGGAGTAATCCTCTATCTATATC 979
QY 763 TTTATGCCAGTTACTGTGTGATCAATAGAACGTATTGACCGATCTCGCTTATTAATC 822
Db 978 AGAAGTTAGATCGCTTGATGATGAGAGACGTCATCTTAAAGAGGGGTTGGCAGCAAA 919
QY 823 CAGAGAAATGCAAAAGGGGAAAAAGGTCTAAACCTTAATGAATATGAAGATAGCT 882
Db 918 TGTGCTGCTCAGCTTAAAGAAAGCAAAAGCGGTGCAAGCAAAAGCAAAAGCTCT 859
QY 883 TCTCTATCTTGAATGATCTCTTATATCTTTTATTAATCTCTGTTATTTGCTATG 942
Db 858 TCGCTCTTCTCCCTCCGCTCCGCTATCTATTAGCATTGCGCGTCTGATCTCTTACA 799
QY 943 GTATTGCAATTAATTCGAAGTACTTTGGAAGTCAAGTCAAGTCAATATCTTA 1002
Db 798 ATTTACTCATTCACCTTTTGAAGTGTGGAACCAAGCTCTCGGATTTATGCTT 739
QY 1003 ATTTGATGACTTATGATGATTCATGAGGAACTTCTCTCTGAGCGGGGTGATTCG 1062
Db 738 CTGAGATAGATTTAATCTTATATGATGATGATTCAGACGCTCACAGGAATGCTCTCAG 679
QY 1063 TACTTATCATGCTATTTGTTGTTGATGATGATGATGATGATGATGATGATGATGATG 1122
Db 678 CTTTGACAGGATTTTTCGCTGCTGAGCAAACTATTCGCGGTGGAGTATGATGATG 619
QY 1123 CCTTACTACTCTGTCATAGGTTCTCTTAACAGATATCGTTTCTT---CGCTTGTGA 1179
Db 618 CCTTGTATCCCTTTTAACCATGTTATTAACAGACCTTTGCTTGTGCGCTATCTATG 559
QY 1180 TCTTTGAACACCAAGCTTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1239
Db 558 CTGTAAAGGAGATGCTATGATTTTGGCGGATCTTAAAGATCTCTCGCTAGTCTTA 499
QY 1240 CTGTGTTCTGAGCTATACAGAAATTTCTTGAATATCAACAAATACGCTCTCTTTG 1299
Db 498 CTGCTTGTAGAGGTGTTCAAGAACGTCCTTTCGAGAGCAATTAAGTTTCACTATTTGG 439

```

QY	1300	ACTCAACTAAGAAATGGCTATATCCCTTGTGACCAAGACAAAAGTCAAGTAAGG	1359
DB	438	ATCAAAACCAAGAGATGGCATTTATCTTTGGAGATGATGAGAATTAACGGTAAGG	379
QY	1360	CTGCTATTATAGTGTGCGCCGCTTGGAAAATCAGAGAGAGCTTTAATCAACAAG	1419
DB	378	CGGCTATTTATGGGATCTCCAGAGTAGGAGGAGGCGCTCTAGTTTACCAAG	319
QY	1420	GTTTGGCTTATCTGTGGAAGTATGAGATGACCCCTATCTTGTGAGTATCTTTC	1479
DB	318	GACTGTGATCATCTTCTGCTGCGGTGACAGCTTAAGTTAGCATTTGCTGTCT	259
QY	1480	TTTTTCATCTATCTATTTGGTTGGTTTGTGCACTAAGTTAAACAACATTTAGCCG	1539
DB	258	TGCTGCTTTAGGAGGTGATCTTTGTGATCGCTGTTAGTAGAGATACACAGCCA	199
QY	1540	AGCTTG 1545	
DB	198	AGACCG 193	
RESULT	29		
LOCUS	TAJ10587	1623 bp	DNA linear BCT 23-MAR-1999
DEFINITION	Chlamydia trachomatis npt2 gene.		
ACCESSION	AJ010587		
VERSION	AJ010587.1	GI:4538861	
KEYWORDS	npt2 gene; nucleoside triphosphate transport protein 2.		
SOURCE	Chlamydia trachomatis.		
ORGANISM	Chlamydia trachomatis.		
REFERENCE	1 (bases 1 to 1623)		
AUTHORS	Tjaden,J., Winkler,H.H., Schwoppe,C., Van Der Laan,M., Mohlmann,T. and Neuhaus,H.E.		
TITLE	Two nucleotide transport proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy		
JOURNAL	J. Bacteriol. 181 (4), 1196-1202 (1999)		
MEDLINE	99138740		
PUBMED	9973346		
REFERENCE	2 (bases 1 to 1623)		
AUTHORS	Neuhaus,E.		
JOURNAL	Direct Submission		
TITLE	Submitted (25-AUG-1998) Neuhaus E., University of Osnabrueck, Biology / Plant Physiology, Barbarastr.11, D-49069 Osnabrueck, FRG		
FEATURES	location/Qualifiers		
SOURCE	1..1623		
gene	/organism="Chlamydia trachomatis"		
CDS	/db_xref="taxon:813"		
	1..1623		
	/gene="npt2"		
	1..1623		
	/function="nucleoside triphosphate uptake"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=1		
	/product="nucleoside triphosphate transport protein 2"		
	/protein_id="CA839535.1"		
	/db_xref="GI:4538862"		
	/translation="MSSEVSKFRGFEPIYSESKFIPLFLAFVGVNVALK		
	TTKDSLIVGRAGAEVIFPLKVGIVGAVITMIGMRSRYSRGVETSLVGFL		
	GFALPFAVYIPIDALHLKLAQLQSLIPGGRGEVAVQVYSLSLYMSLMS		
	VLSLTFMGVANHITTSVREAGFPALINTGNTLSVFAGEVSLMGRSPVAFPMAD		
	PMHEMLNITLLIYIAGCVILYOKLDRLMDERSMLEGLAEMSVOLKKEKRSK		
	AKAKSLRALIRSRVLGIAIVVLSYLVHLEFVMDQVCRITYASVEFNSYMSRI		
	TTLVGISALAGIFPAAGQTLRWGTVGALVPLPILITGLAFGAIYAVFGDMIR		
	GILGISPLVILAMIGVONFSRAIKFTYEDQTEMAFIPLEDEKNKGAIIDGIS		
	RVGSGSLVYOGILIFSSVAASLNAITIVLLALGSMIFVIAMLGREYAKRETVLV		
	RVNASEDDVLEBREASSLVDAESREPEVTTL"		
BASE COUNT	374 a 313 c 410 g 526 t		
ORIGIN			

QY	1300	ACTCAACTAAGAAATGGCTATATCCCTTGTGACCAAGACAAAAGTCAAGTAAGG	1359
DB	438	ATCAAAACCAAGAGATGGCATTTATCTTTGGAGATGATGAGAATTAACGGTAAGG	379
QY	1360	CTGCTATTATAGTGTGCGCCGCTTGGAAAATCAGAGAGAGCTTTAATCAACAAG	1419
DB	378	CGGCTATTTATGGGATCTCCAGAGTAGGAGGAGGCGCTCTAGTTTACCAAG	319
QY	1420	GTTTGGCTTATCTGTGGAAGTATGAGATGACCCCTATCTTGTGAGTATCTTTC	1479
DB	318	GACTGTGATCATCTTCTGCTGCGGTGACAGCTTAAGTTAGCATTTGCTGTCT	259
QY	1480	TTTTTCATCTATCTATTTGGTTGGTTTGTGCACTAAGTTAAACAACATTTAGCCG	1539
DB	258	TGCTGCTTTAGGAGGTGATCTTTGTGATCGCTGTTAGTAGAGATACACAGCCA	199
QY	1540	AGCTTG 1545	
DB	198	AGACCG 193	
Query Match	15.4%;	Score 252.2;	DB 1; Length 1623;
Best Local Similarity	49.3%;	Pred. No. 1.2e-49;	
Matches	758;	Conservative	0; Mismatches 758; Indels 21; Gaps 3;
QY	61	CGAAGAAACCTTTGGAAAATGCGCTCTTCTGTGGCCGATACATCTACAGAC	120
DB	8	CCGAGGTAAATTCCTTTGGAAGTTCAGGGATACATTTTCCCATATACAGTCAGAA	67
QY	121	TAAAGAAATCTGCGCAATGCTCTAATGCTCTGCTATTAATTAATCTATAGCTGT	180
DB	68	TCTCTAATTTATACCTTATCTTTTATAGCTTTTGTAGGTATACATACGCTTTAT	127
QY	181	TACGCGATCAAAAGACACTCTTATTTGTGGAGCTCTGTTCTGTGCGAGAGCAATAC	240
DB	128	TAAAGACCAAGAAAGATCCCTGTTTATAGTGGGCTGAGACAGGGCGGAGATAC	187
QY	241	CTTTCAATCAAGTTTGGCTTGTTCCTGCTGCTATTAATCTTATATGCA	300
DB	188	CTTCTTCAAGCTGCGGGATGTCCTGCGGGCTTATCGTCAACATGATCTATGAT	247
QY	301	AGCTAAGTATATTTTAAGTAAAGACGCTTATTTATGCACTGGGAACGCCCTTTTA	360
DB	248	GGATGAGTCGACGTTATTTCAAGAGTACCGGTATATCTTTAGTCGGCGGTTTTAG	307
QY	361	TTTTCTTGGCCCTGTCCGACTGTAAATTAATCCGCTAGCGATGTTTACATCCACAG	420
DB	308	GCTTTTTCCTGCTGTTCCGACAGATGATTTATCCATAGAGATGCGCTGCACTAAATA	367
QY	421	AATTTGCTGACCGTTTACAGGCCATCTACCTCCAGAGATGCTAGACTCGTGCATCT	480
DB	368	AATTTGGCGCAAAACATACAGTCACTTACCTCCAGGGGAGAGAGGTTTGTAGTAGTG	427
QY	481	TAAAGAACTGACATTTGCTGATTTATGATCTGCTGCTAGCTATGGGAGACGTCATGC	540
DB	428	TTCAATATTGGAGCTATAGCCGTATATGATGATCTGATGATTAAGATTCGTAAGTTT	487
QY	541	TATCTTAATGTTTGGGGATTTGCTAATGAATAACAAAATTCACAGAACAAAGCCTT	600
DB	488	TGCTACTGCTGTTTGGGGGATGAGCAACCAATATACGATTTAGCGAGAGGCGGT	547
QY	601	TCTAGCTCTTTTGGGTATCGGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTG	660
DB	548	TTTACGCTTATTAATCTGATTAACCTCTCTCTTCTTCTGAGAGGTTTCTT	607
QY	661	TTTGG---GCTTCAAGTTGAGAGCTCGTTCGTAAGGTGTAGATCTTG-----	710
DB	608	TGTGGCTGGTAGAAGTCTGTAATTTGCTTCCCTATGCGGTAGATCTTGCAATGAGA	667
QY	711	-----GGAATTTCTTACGCTTTTTCATGCGTATGACTAATTTATCTTGACTTGTTC	762
DB	668	TGTTGCTCAATATCACTACTTATCTGATGCGCGGAGATTAATCCATCTATATAC	727
QY	763	TTATGCCAGTTACTGCTGATCAATTAAGACGTATTGACCGATCTCGCTTCTTAATC	822
DB	728	AGAACTTGAATCGCTTGAATGATGAGAGCTCAATTTTGAAGAGGGGTTTGGCAGAGAA	787
QY	823	CAGAAGAAATGCAAAAGGGAAGAAAGGTGCTAAACCTAAATTAATTAATTAAGATAC	882
DB	788	TGTGCTGCTGCTGCTTGAAGAAAGGAAAGCGTCCGAGGCAAAAGCAAAAGTCTCT	847
QY	883	TGCTCTATCTGATGATGATCTCTTATTTCTTTTATTAATCTCTTGGTTATGCGCTATG	942
DB	848	TGCGCTCTCTCTCGCGTCCCGCTATCTATTAAGCATGCGCGTCTGCTACTCTTACA	907
QY	943	GTAATTCATTAATTAATCAAGTGAATTTGAAAAGTCAGCTGAACATGCAATATCTTA	1002
DB	908	ATTATGATTAATCACTTTTGAAGTCGTTTGAAGCAACAGTCTGTGGAATTAATGCTT	967
QY	1003	ATATGAAGACATATGATGATCAATGGGAACTTCTCTTGAGCGTGGCGATGATCCG	1062
DB	968	CTCGATGATTAATTTATTTATGATGATGATGATGATGATGATGATGATGATGATG	1027

QY	1063	TACTATCATCTATTTGCTGTGGTAAAGCGCATTCGTAATTTGGATGCTTAACCTAGAG	1122
Db	1028	CTTTGGCAGGAGATTTTGGCTCTCTGGACAACTATTTCCCGTTGGGGATGACTGTGGCG	1087
QY	1123	CCCTAGTCACCTCTGTCATGGTTCTCTCAACAGATATCGTTTCTT---CCACTCTTGTTA	1179
Db	1088	CTTTGGTACCCCTTTTAACCAATTAATGATTAGATACAGAGACCTTGTCTTGGCGCTATCATAG	1147
QY	1180	TCCTTACAAACCAACCTTCTGGGCTGGTGCCTATGTCGTGTAACACCTCTTCATGCTAG	1239
Db	1148	CTGTAAAGGGGATGCTATGATTTTGGCGGGATCTTATGGGATCTCTCCCTAGTCTTAA	1207
QY	1240	CTGGGCTGTGGGAGCTATACGAATATTTCTTGGAAATTCACAAATATAGCTCTCTTGG	1299
Db	1208	CTTGCTGTATAGGAGTGTTCAGAACGCTTCTCTGAGAGCAATTAAGTTCACTATATTCG	1267
QY	1300	ACTCACTAAAGAAATGGCCCTATATCCCTCTTGACCAAGACCAAAATGCTCAAGGTAAG	1359
Db	1268	ATCAACAACAGAGATGGCGCTTATTCCTTTGGAAAGATGATGACAAAGATTAACGTAAG	1327
QY	1360	CTGCTATTTGATGTAGTGGCCCGCCGCTTCGGAAATTCAGAGAGCTTTAATCCAAACAG	1419
Db	1328	CGGCTATTTAGATGGGGTGATCTCCAAAGTAGGAATGACAGGGGCTTTAGTTAACCAAG	1387
QY	1420	GTTTGCCTGCTATATCTGTGGAGATATGGAGCTATGACCCCTATCTTCTGACTGATCTTC	1479
Db	1388	GACTGCTGATCATCTTCTCGTCCGTTGGACATGCTTAAATGCGATTAACGATTTGTTGT	1447
QY	1480	TTTTCATCATTTGCTATTTGGTTGGTTCTGCAACATGAATTAACAACTATTTTAGCGC	1539
Db	1448	TGCTGCTTTAGGAGATTTGGATCTTTTGATGATCGCTTGGTTAGTAGAGATACACAGCCA	1507
QY	1540	AGTCTGCTCTTAAAGAACAGAGATGGCTCAAGAGA	1576
Db	1508	AGACGAGACACTCTTGTAGATCAATGCTCTGAAGA	1544
RESULT 30			
RPXX02			
LOCUS			
DEFINITION	RPXX02	312430 bp	DNA linear BCT 12-NOV-1998
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
COMMENT			
FEATURES			
Source			
gene			
CDS			

gene
CDS
/gene="RP226"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="CA14689.1"
/db_xref="GI:3860789"
/translation="MLKCLDILPLINLLVTVOGVSPPLPPSLIIVAPDTKNI
STNSISFEFKQPFSSKÖKKRNISSCHÖDÖRÖKAIHÖSSÖQIDÖRELNEDÖSEPTI
DÖSTÖTILPSVANSYIDÖKAEYENSTNIAASNTÖDIDVÖKÖÖDPÖSEASÖPIDIGMTK
FTSÄNHNEÖEAVSNDKRETNILITPTNVSÖPVSISPTADÖVNVSPÖSÖVOYI
KPTNLSIRNÖIPLNHTDNLKVEKNLESTISMTÖTPTIMVÖVSPÖIÖTÖTLNIT
VPTATÖHVÖSTÖTVMNSHNSÖAQÖPTISÖTINPEVÖSÖVLÖATÖESSÖVÖINNSÖÖF
VÖSESÖTKÖMÖVÖPTIMÖVLVYÖDPNKSÖQSPÖALEÖKNNÖDÖIÖNNÖÖSHVSÖSSNV
TÖÖKÖDÖNNVÖNÖATÖESTÖREÖFVNÖTÖMÖLÖPÖDDÖIVÖGÖKLEÖATÖEDÖMÖBYÖIKLE
ÖÖKREÖVMÖNÖÄRÖKRLVÖSEÖFIÖYDDÖIKNNÖDÖIÖANÖSYÖSVÖNÖÄRÖKÖMÖFLÖR
ÖLÖDDVÖYÖPLÖÖKÖNNSÖTÖGÖELLÖSSÖYÖNGÖNYÖVLÖAFÖLVÖIGÖIKÖTÖSVÖLÖKNDÖCKYÖPLDÖIAL
ÖKNNÖIVÖCMÖKÖKÖKÖVÖN"
2025, 4013
/gene="RP227"
2025, 4013
/gene="RP227"
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit B (gyrB)"
/protein_id="CA14690.1"
/db_xref="GI:3860790"
/translation="MSDLFSLNKKRNKIVYTNYSANDIEVLÖDLÖEÖFRKÖPGMYÖIG
TÖDÖSNMHÖLVÖSEÖVLÖDNÖMÖEÖVÖAGÖFÖASÖIÖMÖKMHÖDSÖITÖPDÖNGÖCIÖPDÖNHÖKÖPD
KÖSÖLEÖVÖLÖITÖTÖHÖSGÖKÖSFÖNNÖYHÖTÖSGÖLÖHGÖIGÖSVÖNÖLÖSKÖHKÖIVYÖKÖGÖKÖISÖYÖ
SKÖGÖKÖTÖLÖJÖSÖEÖSKÖRLÖGÖTSÖINÖTPÖDPÖELÖFSÖKÖLÖHÖFNÖKÖIÖELÖARSÖAYÖIÖXÖRYS
IÖMÖEÖVÖEÖSPÖDÖFRKÖALÖINÖPÖNGÖLÖTÖLSÖKÖISÖDÖNLVÖIÖELFSÖGÖNÖESTÖVDÖIKLE
VÖMÖÖKÖNNDÖTÖSVÖFÖQÖCYÖNÖVÖPTÖQÖGÖTHÖÖGÖKSÖLÖRLÖKÖYÖSEMÖÖGÖKÖKÖANÖLÖI
IÖDLÖETÖSÖIVÖSÖIFÖIÖVÖEÖSPÖQÖÖTÖKÖERÖLVÖSÖGÖSÖLEVENÖTÖIKÖHÖDÖHÖFÖLÖSSÖDÖVLÖATÖH
LÖEHÖYÖIÖÄÖERRÖNKKÖNÖERNÖRSÖKÖSTÖYÖKÖLÖPKÖLÖADÖCTÖSÖEÖGÖELÖTÖIÖEGÖDSÖG
GÖSÖKÖÄRÖNÖRÖÖVÖYÖPLÖMGÖKÖVLÖNÖASÖTÖLEKÖIÖNNÖÖIÖDLÖIÖALÖGÖSLÖKÖKÖNNÖL
RÖEKÖIÖTÖNÖÖVÖGÖAHÖSÖLÖMTÖPÖFLÖRÖKÖLÖVÖEÖSHÖLÖLÖRÖPÖLÖRÖLÖGÖKÖKÖIÖYÖA
CÖEÖEKÖIÖLÖTÖIKÖLSÖKÖSÖKÖLÖEÖVÖRÖGÖLÖGÖMÖPÖLÖKÖETÖMÖEÖKÖSLÖKÖVLÖEÖVÖQ
NÖDKÖIÖVDÖLÖMÖKÖKÖEÖKÖRÖQÖIÖYÖÖLÖVÖKÖMÖQÖIÖNKKÖDI"
4122, 5447
/gene="RP228"
4122, 5447
/gene="RP228"
/codon_start=1
/transl_table=1
/product="tail-specific protease precursor (cup)"
/protein_id="CA14691.1"
/db_xref="GI:3860791"
/translation="MLRFLIÖLPLSLNÖCÖALEÖGÖETÖENKÖISNÖEÖKÖQÖDÖVÖERÖLEÖK
DÖYÖVÖDÖÖKÖMÖIDEÖAINÖGMLÖSLDÖPHÖSNÖYÖTÖDEÖLÖDIÖFTÖGÖEÖFGÖIÖVEÖIMÖYÖSG
ÖAÖIÖSSÖIÖDDÖLPÖRFÖAGÖLÖGÖDÖIVÖGÖVÖDNÖÖKÖLÖKÖEMÖKÖGÖPÖKÖVÖLÖIÖIK
ÖEÖEÖPÖÖLÖEÖLÖTÖREÖIVÖIKÖPÖIKÖHÖLEKNNÖIÖAYÖIÖTÖTFÖNÖSTÖISÖELÖKÖAÖKÖKLÖTÖSK
DÖNLÖKIÖIÖDLÖNNÖMGÖIÖDÖNÖIÖVÖDÖYÖEÖDSÖGÖVÖVÖTÖKÖGÖTSÖNSÖETÖKANÖEÖSÖLÖK
KÖVÖMÖKÖIÖIÖDÖLÖNNSÖSÖSÖSETÖVÖAÖLÖDÖHÖRÖIÖLÖIGÖKÖSÖSÖVÖALÖÖTÖNÖSÖRÖAVÖKÖTI
SKÖYTÖPÖSÖRSÖIÖÖEÖIGÖEPÖDÖIÖLÖPÖKÖVÖPEÖVKÖIDÖKÖRÖFÖSÖSSÖKÖNTÖLÖKÖDÖNÖKÖKÖDS
NÖKÖTÖYÖTÖKÖEÖEÖSELÖSEÖLÖYÖKÖDÖYÖFÖRÖVÖDÖYÖTÖGÖLÖIÖNÖTÖLÖÖKÖKÖ"
5653, 7146
/gene="RP229"
5653, 7146
/gene="RP229"
/codon_start=1
/transl_table=1
/product="histidine kinase sensor protein (barA)"
/protein_id="CA14692.1"
/db_xref="GI:3860792"
/translation="MNHKÖYÖVÖRÖSLÖTÖIÖPLÖNÖVÖIÖNNÖLÖFRÖYÖMÖIKÖMÖIÖKÖVÖALEÖHT
KÖIÖVÖLYÖDNÖIÖNHNÖKÖNÖYÖSLÖKÖHÖFÖDÖYÖLÖKLÖÖDÖDÖFÖNÖVÖÖTÖKÖMÖFÖNÖIÖNÖISÖLÖD
KÖNÖKÖTÖSNÖMLÖHÖSVÖDÖYÖKÖDSÖLÖEÖIVÖYÖIKÖDÖIFÖLÖKÖTÖSÖAPÖLÖADÖKÖGÖITÖSHÖL
LPÖKÖYÖENÖESÖLÖTÖIÖEÖHSÖFÖYÖSÖIÖPVÖIDÖHNÖDÖVÖDÖVÖEÖIÖNÖTÖNÖISÖÖKÖNÖISÖLÖÖK
VÖITÖPÖIÖIÖEÖCTÖIÖISÖNTÖYÖARÖIÖEÖDÖLETÖRNLÖKÖIÖEÖKÖVÖERTÖSSÖNÖKÖTFÖPÖNI
SÖHEÖTÖPÖIÖLÖIÖGÖSEÖIÖMSÖRÖDÖEÖKÖKÖYÖKÖIÖKÖHÖGÖKÖHÖLÖSÖIÖNDÖIÖDLÖSÖKÖSD
LÖKÖVÖDÖIDÖDÖNNÖLÖSSÖLÖILÖKÖPÖRÖYÖÖEÖVÖLLÖSÖRÖLÖKÖEÖHÖVÖIÖNÖADÖKÖRÖLÖKÖVÖLÖNL
NÖVÖEÖGÖLÖSÖGÖSVÖTÖIÖEÖKÖDÖLÖVÖYÖIKÖIÖDÖIGÖIÖEÖKDÖIÖPTÖSÖFÖQÖIÖDÖSELÖSÖR
YÖEÖGÖLÖIÖPLÖTKÖIÖVÖLÖMÖNÖKÖPÖLÖÖKÖKÖNGÖTÖVÖYÖTÖFRÖYÖDÖSÖIÖEÖ"

gene 7374. .8126
/gene="RP230"
CDS 7374. .8126
/gene="RP230"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAI14693.1"
/db_xref="GI:3860793"
/translation="MKLIITLLPFLSFLSLGKSEMPQKPLKAVNDFENRLDEOE
QIRLIGVEVLLQKIDILKONLMUNDOEVI EVLFEFDDPKOD I FPIALLEGHND
VSKRFEEVKKDLPYKQAVDLALAAVYKKNLLEAKDKRKNLTKYKRNLSI.SNMFYEW
ACPEFKQDYNCAIINYLCKYKESPKGAKSSDGLKLALSLGELKKMOEACNIILAKLE
KEPFINRISYSKMKTEDAKIKFGCAIKNNK1"
8123. .8785
/gene="RP231"
CDS 8123. .8785
/gene="RP231"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAI14694.1"
/db_xref="GI:3860794"
/translation="MTDLDEVLDQNEERKLPEPKLVIPIIIISVITVTIWIINK
NKDNRIKNNQKGDILVKTIGLDTTKDKKALAFNTLENLITSSNRKIKETALSOVMI
YDSEKRYGKAKNLKNIIDNEKYSISIAVARIAMCCIVLDDQSILNTQDKRKLORYL
FDDSKRFGPWAATATIIKALILDIKHNKTKQAEKNILKLASNNSDCLKDAQALLVSL
SK"
gene 8812. .10059
/gene="RP232"
CDS 8812. .10059
/gene="RP232"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAI14695.1"
/db_xref="GI:3860795"
/translation="MKKKLALLLPILISGNGLSKSVKNIVDTPLKVIQTNEPIY
LDSNRKIYIPVYNNMLKNQYSPAKSKMAEVPFGDMITADIRANISAFIEKNKIL
WSYLSNRKKNQYIGGILHNHGKLIYVYGARLLVIDAKSGEITREKLEDIRIKR
IALNHDITLNOTISNOTIALDSEPLKTLMDHSEIAELISTGYSMPIVOHNVIVYVYN
TGQVALNIKNGEYKWNFEFTNLNDHTIIPNFTSSILCTPYHDSMNIYATGGLKL
KLNLITGVLWQINADIDQSMGLNSLFIINMARQIALNPETGKVFADLVENRN
DKRLKSTPLFVPEVGVDNNQORSNLVSDGILVNFINDSGLMKMNPDIILINIIY
YGLRSNNLTYESSTDQVIFGSI1"
gene 10279. .10746
/gene="RP233"
CDS 10279. .10746
/gene="RP233"
/codon_start=1
/transl_table=11
/product="50S RIBOSOMAL PROTEIN L13 (rplM)"
/protein_id="CAI14696.1"
/db_xref="GI:3860796"
/translation="WKITYSAKPSIEIKKMWIDAKNVYLGRGLASRAVIMLGRKHPSF
TPHLDGCGNIITIIINAEHIKLTGKLNPDKGVVYVRHGFPGIGIKDTAGKILSGYPEE
RVKIAVCAVRMTIRTNLVAQKQMSNLVYVANCBEHPMAOQPTIYDPASEMPKNKK"
gene 10753. .11238
/gene="RP234"
CDS 10753. .11238
/gene="RP234"
/codon_start=1
/transl_table=11
/product="30S RIBOSOMAL PROTEIN S9 (rpsI)"
/protein_id="CAI14697.1"
/db_xref="GI:3860797"
/translation="MTELKIKTEKVVVKQLTKESSKSVLKIPKEIDSVSRYFATGRRK
NALIAVLMVKYEGKGVIVNNKILNOVFSEYFVKTIILOPILTKTIIDQYDVCYKGGG
ISGQGAIIHGHSIALDKSAPCPHAILRKSGLLLRDSVVERKKYGGRRKAKRKTQFCK
R"
gene 11279. .11352
/gene="rRNA Met (CAT)"
CDS 11279. .11352

Query Match	Best Local Similarity	Score	232.8	DB 1	Length	312430			
Matches	774	Conservative	0	Mismatches	737	Indels	39	Gaps	6
gene	/gene="tRNA Met (CAT)"	complement(12284 . .12769)							
CDS	/gene="RP236"	complement(12284 . .12769)							
gene	/gene="RP236"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							
gene	/gene="RP237"	complement(12798 . .14150)							
CDS	/gene="RP237"	complement(12798 . .14150)							

QY 730 TGATGGCTATGACTATGTATCTGACCTGTTCTTATGCGCAGTACTGCTGATCAATA 789
 Db 309771 TCATGCTTACGACCATTCATATTAATTTATCTAGTCAGGAATATTCGTATGTTTAT 309830
 QY 790 AGACGATTTGACCGACTCCTCGCTCTATATCCAGAGAATAATGCAAAAGGGGAAAAAG 849
 Db 309831 TTAGCAATATATAAATTAATTTATTAACATAATCTATAAATTTTACATGTAATAAAG 309890
 QY 850 GTGCT-----AACTATAATGAAATATGAAGATAGCTCTCTCTATCTTGATAGT 900
 Db 309891 TTGGCGCTAAACAAAACAAAACCTGCAATTAATGAAGATTAATAAATTAATTAATTCATT 309950
 QY 901 CTCCTTATATCTTTTATTAATCTCTTGCTATGCTGATGCTGATGCTTATGCTTAACTTAA 960
 Db 309951 CAAATATATATAGCTGATATGATATATATATATCTGTTATGATATACATAATATATAG 310010
 QY 961 TCGAAGTACTTGGAAAGTCAAGTGAACCTGCAATATCCATATATGAAATGACTATAGT 1020
 Db 310011 TTGAAGACCTTGGAAAGGAAATAAAGAAATTAACATCAATATGATGATATGTTA 310070
 QY 1021 AGTTCATGGGGAACCTTCTCTCTGACCTGCGCTAGTATCCCTACTTATCATGCTATTTG 1080
 Db 310071 ATTTATGCGGCAATGTTTATATATTTGATGCGGATCTCATGCTGCTACTTTCATGATATA 310129
 QY 1081 TTGGTGTACAGCTCATTCCTTAATTTGATGCTGATGCTGATGCTGATGCTGATGCTGCTG 1140
 Db 310130 --GGATGATATATTTCTTGAAGGCTTGCTGCTATTTCTCATATATTAACCTCTATTA 310187
 QY 1141 TGGTCTCTCTACAGCTATGCTTCTCTGCTCTGCTTATCTTATGAAACAGCTTCTG 1200
 Db 310188 TGTATATATATACAGGCTTCATGTTTATTTATATATATTTTATGAAATAATAGTA 310247
 QY 1201 GGCTGCTGCTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 Db 310248 CATGTTTGGTGAATTT-----TAATCTCTATGATGCTGCTGCTGCTGCTGCTGCTGCTG 310301
 QY 1261 AGAATATCTTCTGAAATCCAAATATACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 Db 310302 AGAATATCTTCTGAAATCCAAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 310361
 QY 1321 ATATCCCTCTGACCAAGAGCAAAAGTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 Db 310362 ATATCCCTCTGACCAAGAGCAAAAGTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 310421
 QY 1381 CCGGCTTGGGAAATCAGAGAGAGCTTATATCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1436
 Db 310422 CGAAATTTGGTAAATCATTGGAGCAATTTATCCAGCTTGTGATATTTATTTATTTCCAA 310481
 QY 1437 --GGAATATTTGAGCTATGACCCCTTATCTGAGTATCTCTGCTGCTGCTGCTGCTGCTGCTG 1494
 Db 310482 CGGCTACCTTTGATCTATATATATATTTTACTAGTATTTTATAGTATGATGATGATTT 310541
 QY 1495 TTTGGTGGTTTCTGCAACTAAGTTAAACAACTATTTCTTACGCGAGTCT 1514
 Db 310542 TATGATTTGCAATATTTAATAATTAAGAAATATATTAAGCTGCT 310591
 RESULT 31
 AE002346/c 12084 bp DNA linear BCT 26-MAY-2000
 LOCUS Chlamydia muridarum, section 74 of 85 of the complete genome.
 DEFINITION AE002346 AE002160
 ACCESSION
 VERSION AE002346.2 GI:8163321
 KEYWORDS
 SOURCE Chlamydia muridarum
 ORGANISM Chlamydia muridarum
 Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
 REFERENCE
 AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
 White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Ulteback,T.,
 Berry,K., Bass,S., Linner,K., Weidman,J., Khouli,H., Craven,B.,
 Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J.,
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

TITLE Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 JOURNAL pneumoniae AR39
 MEDLINE Nucleic Acids Res. 28 (6), 1397-1406 (2000)
 PUBMED 20150255
 REFERENCE 10684935
 AUTHORS 2 (bases 1 to 12084)
 Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
 White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Ulteback,T.,
 Berry,K., Bass,S., Linner,K., Weidman,J., Khouli,H., Craven,B.,
 Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J.,
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
 TITLE Submitted
 JOURNAL Direct Submission
 COMMENT Medical Center Dr. Rockville, MD 20850, USA
 FEATURES On Jun 1, 2000 this sequence version replaced gi:7190805.
 source Location/Qualifiers
 1..12084
 /organism="Chlamydia muridarum"
 /db_xref="taxon:83560"
 /note="synonym: Chlamydia trachomatis Mopn"
 complement(96..713)
 /gene="TC0779"
 complement(96..713)
 /gene="TC0779"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF3582.1"
 /db_xref="GI:7190805"
 /translation="MGGMLDLKISVTGDPSSGCTEACQVEELGAPYSADKXSHF
 LVPYISGQAVVDLGPETIVENTLNRAIAKKVFNGLDLSLEKILHPGCRVEE
 NYTOYQAKKSLFVENVPLVEIOYAPMPFRVILISADTEIRKRFLLKTKGSDTSF
 DLRCARFSSLEKILRADVYVLENNGTKEEFHKKVQCFKALKGT"
 complement(698..3298)
 /gene="TC0780"
 complement(698..3298)
 /note="TC0780"
 /note="similar to GB:J01663 SP:P00582 GB:J01664 GB:V00317
 PID:147312; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="DNA polymerase I"
 /protein_id="AAF3583.1"
 /db_xref="GI:7190807"
 /translation="MKKLEILDASGFIFRAYFALPEMKGNGSTQAVFGFIRSLDKL
 VNDLSEYVAIVFDGPNKOSROELIADYKKNRDKFEDLPEQIALVQYCELIIGIPC
 LEKESEVDVIVASVAKAVAGEFVCITADKDLQVNSOVSVFNKROEIKYNE
 VILQGVAPDQIADYLAIVGSSSDNIPVQCGCKRAQAFLEKROSVEDVIVANTORLS
 SKNKMITEQIETLLSKRLATLHTDLMLPLKTEBMSFSOTADVAQOLNARYLQNGK
 ALVKHAPSETSILOVVPKPSVLDVNLKNGKRGVGCVAATYGEHLISLHGVALS
 GDEVFYIEISGDDIELILKAFEDATTKFPGYRKRBNHAKLNGKIDVYADLVLA
 OHLVNGAKISFOTLLVESGHIQEAFAFAKAGASLSPVONLPONPAQYFGIYVSRIP
 SIKNTLFEKLEKIKSIFKNEVDLEPLIEILPAMCAGMPLDSEGLVILDRDISELET
 YTOETIYDAGEPINKSPKQSLSDVLYIHLGILKPDNKAETAEVLEALEGHEITSKI
 LAFRATERMLSTIVKALPRQIDLGHVHPFNQVGTGRLSCDPPNLQNIPIKTER
 GKSLEAFRTKENDYFLAADYQVLEFLAHLSDDETLKRAPDSGEDIHFTASQV
 NVPLDQVTERQYQAKAVNGLAYQCOQVGLSKNLKISVNAQGLIDVYFARYQASV
 FITHITBOASKEOKVTMLGRLITSDRESSPGAAASGLAVNTRIGQSAEILIKLA
 MLNISRLEVSRLGKSRLLQIHDELLEFVPRDELEEMKVLVQEKMSMSLSVPLVAV
 VLIGNNAEC"
 complement(3312..4310)
 /gene="TC0781"
 complement(3312..4310)
 /gene="TC0781"
 /note="similar to GP:147868; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="protease IV, putative"
 /protein_id="AAF3584.1"
 /db_xref="GI:7190808"

QY	425	TGCGACCGTTTACAGGCCATTCCTACCTCGAGATTCGAGGACGTCGCTGGCATCTTAAG	484
Db	5712	GCGCGCAAAACGTGACACTATTCCTGCTCGAGAGCTCGAGGTTTCGTAGTATGTTTCA	56533
QY	485	AAACTGACATTTGCTGCATATTTATGTACTGCTGACATATGGGACGCGATCTCATC	544
Db	5652	GTACTGAGACACAGCTTGTACTACCTAATGTCTAGCTATATGAGATTCGCTTGTTTATC	55933
QY	545	TCCTATGTTCTGGGGATTGCTATAGTAATTTACAAAATTCACGAGCAAGACGTTTCTA	604
Db	5592	CACCTTATTTTGGGGGATGCTCAACCAACATTCAGATGTCGTGAAGCAGGCGCTTTTA	55333
QY	605	CGCTCTTTTGGGATCGGACCTAATATTCTTTACTACCTTGGTGCATTTGTTG	664
Db	5532	CGCTCTATTAATGTTGGATTAAATGTCTCTTATTTGTTGGGAGAAATTTCTCTCG	54733
QY	665	--GCGTTCAAGTTGAGACCTCCGTTTCTGAAAGTGTAGATTCCTGGGGAAATTTCTTT	721
Db	5472	GCTGGGTAAACATACGTTAATTCCTTCCTCTATGCTGTGGATGCTTGACGCGGATAT	54133
QY	722	ACGCTTTTGTATGGCTATGACATTTGATCTGTGACTGTTCTTATGCGCAGTACTGCTG	781
Db	5412	ACCTAATATTACCTGCTGTGATTTGCTGCTGAGGAGATTTGATTCCTTTATACCGAA	53533
QY	782	GATCAATATAGACGTATTGACCGCATCTCGCTTCTATTAATCAGAAAGAAATTCAGAAAGG	841
Db	5352	ACTGACCATTTTAACAGAAAGCGCTCTGATTAGAGACGATTTGTGAGAAATGTC	52933
QY	842	GAAMAAAGTGCTAAACCTTAAATGSAATATG-----AAAGATAGCTTCT	886
Db	5292	GGTGTCTCAGCTCAGACAGAGAAAGCGGCTTAAGCAAAAGCAAAAGTCTTCTGTC	52333
QY	887	CTATCTTGATAGATCTCTCTTAATTTCTTTTATTAATCTCTTGTTATTTGCTATGAT	946
Db	5232	TGTTCTGTCGCTCCGCTCCGCTAATTTAAGGAATGCTGTGTTGTTGCTTCGCTTACACT	51733
QY	947	TTTGATTACTTAATGCAATGACTGTTGGAAAATGCAAGTCAACATCTCTAATAT	1006
Db	5172	AGCTATCCATCTGTAAACATTTTGTGGAGAGCAAGTCTCAAACTATATCTCTCG	51133
QY	1007	GAATGACTATAGTAGTTCATGTTGGGAACTTCTCTGACTGGGCGTAGTATCCGCTACT	1066
Db	5112	AGTAGAATTTAATTTCTTCACTATGATGAGATTAACGTTTCACTGGGATATGTCCTCT	50533
QY	1067	TATCATCTATTTGTTGGTGAAGCTCATTCGTAAATTTGGATGTTAACTGAGACCT	1126
Db	5052	GCGAGGGGTTTTGCTCAGCGGCAAGATATCCGCGCGTGGGGTGAGATGTTGGAGCTT	49933
QY	1127	AGTACCTCTCTCATGTTCTCTTAACAGGTATCGTTTCTT---CGCTTGTATATCT	11833
Db	4992	GATTAATCCCTTAACGATGTTGTAATTCAGGGGGGCTCTCTTTTGGTGCATTAATGCTGT	49333
QY	1184	TAGAAACCAAGCTTCTGGGCTGGCTCATGTTGCTACAACCTCTCTCATGCTAGCTGT	12433
Db	4932	AAAAGGCGATGCGATGATTTTAAAGCGGATTTCTTGATTTTCCCGCATGTTCTACAGC	48733
QY	1244	GGTGTGCGAGCTATACAGATATTTCTTTGGAATTCACAAATACGCTCTTGAATC	13033
Db	4872	TTGGCTGGAGGTGTGAGAACGCTCTTCTTCAAGCAATTAATTTACTTACTTGTATCA	48133
QY	1304	AACTAAAGAAATGGCTATATCCCTCTTGACCAAGAGCAAAAGTCAAGGTAAAGCTGC	13633
Db	4812	AACTAAAGAAATGGCATTTATTCCTTTGGAAATGATGAGAAATTTATGTAAGGCGCG	47533
QY	1364	TATTAATGATGTTGGCGCGGCTTGCGAAATCAGGAGGAGCTTAAATCCAAAGGTTT	14233
Db	4752	TATTAATGAGATGATTTCTAGGGTGTGTAAGTGGGGGATCTTTGGTTTACCAAGCAT	46933
QY	1424	GCTGTTATCTGTGGAAGTATTTGGACTATGACCCCTATCTTGGAGATGATCTCTTTT	14833
Db	4692	ATTGATCATTTTCTCTCTGTGTGACAGATTGCAATGAGATGCGATTACTATTTGCTTAAT	46333
QY	1484	CATCATTTGCTATTTGGTTGCTTCTGCAACTAAGTTAAACAAACATATCTTAAGCGCATC	15433

Db	4632	ACCTTTAGAGCATGCGATTGGGTAGTTCCTGGTTAGCTAAAGATATTCGTTAGAAC	4573
QY	1544	TGCTCTTAAGACAACAAGATGGCTGCACAGAGATTCACGCT	1583
Db	4572	CGCTGCTTTAGCTAAGCGCAGGCGCTGCGTGAAGACGCCGCT	4533
RESULT 32			
LOCUS	AE008630	10893 bp	DNA linear BCT 14-SEP-2001
DEFINITION	Rickettsia conorii Malish 7, section 62 of 114 of the complete genome.		
ACCESSION	AE008630		
VERSION	AE006914		
KEYWORDS	AE008630.1	GI:15619810	
SOURCE			
ORGANISM	Rickettsia conorii.		
REFERENCE	Rickettsia conorii.		
AUTHORS	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group. 1 (sites)		
TITLE	Ogata, H., Audic, S., Barbe, V., Arliaguenave, F., Fournier, P.E., Raoult, D. and Claverie, J.M.		
JOURNAL	Selfish DNA in protein-coding genes of Rickettsia		
MEDLINE	Science 290 (5490), 347-350 (2000)		
PUBMED	20485642		
REFERENCE	11030655		
AUTHORS	2 (sites)		
TITLE	Ogata, H., Audic, S. and Claverie, J.-M.		
JOURNAL	Selfish DNA and the origin of genes		
REFERENCE	Science 291 (5502), 252-253 (2001)		
AUTHORS	3 (bases 1 to 10893)		
TITLE	Ogata, H., Audic, S., Rensato-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.		
JOURNAL	Mechanisms of evolution in Rickettsia conorii and R. prowazekii		
MEDLINE	Science 293 (5537), 2093-2098 (2001)		
PUBMED	21442074		
REFERENCE	11557893		
AUTHORS	4 (bases 1 to 10893)		
TITLE	Ogata, H., Audic, S., Rensato-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.		
JOURNAL	Direct Submission		
COMMENT	Submitted (26-Apr-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France		
FEATURES	A public version of R. conorii genome database is accessible at http://igs-server.cmr-mrs.fr/ . The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.		
SOURCE	Location/Qualifiers		
gene	1..10893		
CDS	/organism="Rickettsia conorii"		
	/strain="Malish 7"		
	/db_xref="taxon:781"		
	complement(326..1579)		
	/gene="ampG1"		
	/note="RC0718"		
	complement(326..1579)		
	/gene="ampG1"		
	/codon_start=1		
	/transl_table=1		
	/product="ampG protein"		
	/protein_id="AAL03256.1"		
	/db_xref="gi:15619811"		
	/translation="MLNSRLCIIMLFGLISGFNIMLTGNTLANYLAKEDIALOTIGITLSFTLPISINFLAPIFDVAOIVYLNKIRHSLSICSSALIFLIYFSPDPSTNLILPAPALISFSPSAODITLSALREIVKRSIGTSQIYFGRVCMLAGSAAIYLSITLTFNEITKIRGLVNTIYILILVIGIKCYDLDENHIQIKDKINOKKKRIINFTINAKRPIGSVYFIILILIFLVLRPLNLNVMINPFLHLEVDADIEASVGRFICGVNVAIGGLVGVIMKHKNLSIFFGIHLIGHILFFELINCKNSLLEFIIITG		

repeat_region
repeat_region
gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene
CDS

IASISGCGMTTAVIAFISALOCGRRAQYPSLSSMGLSISIPFIISGYVWVFQMG
NPFLFTTTITTSLLILKTRKLQ"
1613..1693
/note="REP04, repeated element"
complement(1695..1735)
/note="REP04, repeated element"
complement(1820..1999)
/gene="RC0719"
complement(1820..1999)
/gene="RC0719"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03257.1"
/db_xref="GI:15619812"
/translation="MRRLYEINSLAHSMKOGIEGRKKETIMAKENKKGLSETTIT
MITTLKDKDIEKLK"
2336..2536
/gene="RC0720"
2336..2536
/gene="RC0720"
/codon_start=1
/transl_table=11
/product="similarity to lipopolysaccharide
1,2-glucosyltransferase (rfaj)"
/protein_id="AAL03258.1"
/db_xref="GI:15619813"
/translation="MFKCKRNIITHSYNSRVFLNLKNMKODKDNMLETLOKSNC
DESPMODLNLDIDELFYIF"
2508..2702
/gene="RC0721"
2508..2702
/gene="RC0721"
/codon_start=1
/transl_table=11
/product="similarity to lipopolysaccharide
1,2-glucosyltransferase (rfaj)"
/protein_id="AAL03259.1"
/db_xref="GI:15619814"
/translation="MRMNCFYFEFTQSPFSYFILHVAGPKPWTTDQDELMTNQDLK
DKITKYVMRYKEITPMSSIN"
2771..4276
/gene="tlc3"
/note="RC0722"
2771..4276
/gene="tlc3"
/codon_start=1
/transl_table=11
/product="ADP-ATP carrier protein"
/protein_id="AAL03260.1"
/db_xref="GI:15619815"
/translation="MLPPKIFFEKEYEKIIMPDIERREKLFIPTMALMLCILFNFGALR
SKLSDIYVPMDAGEIISFLKLMVLVPSGVIFTVVYVKLSNLIINFETIFYIVGSFLF
FLEDFIAYPNDDIYHPDENMIKLIASYPNFKMFIRKGSOWSVALMYFAELMSAVV
INLMWGNAINEDPTRSKARFPYLVGMVGNGILIIASVLAFFSGGDVIDSTLPSS
FNSSGNAMIMPDISTSVTAGTIAMLFRINPELFDSTINVDARKYAKTKTIS
VIESKLVIHSHKYIGRIALLITCYGLINIVEGMKAKIKELHNTIIDIVNMKGREVI
NWGISVPIIIGSNILRKLGMLISALTPTMISITGLMEFITLTIEEGCEGDGN
LMAVAIIGAIIINSLKSSKYSLPDSTRKEMAYIPLSLELRKCAAAVEVIGTFKGLJ
GAFIOSLFIIFIPTFDSIIYYLTIFIVAMSIMINNVIKNEVEVELK"
4296..6098
/gene="RC0723"
4296..6098
/gene="RC0723"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03261.1"
/db_xref="GI:15619816"
/translation="MLRRYHRVYASSRLKGNLSKSIIPCONDERTPYPNKRSHYN
TDLKRLKILITFTFTIILA YSGLWFTMESLSHSINOKYAQTNLNIEADNNPQO
OYLKFSVVOVGPPEFKGISVINMOESINRVTETPTNIGIDLKQLFTFHSGE
AFGKKRPVGFVGVYENMCILSTKIPLNKLFIEIVRFKKDLFEVINILENKFTSG

	repeatL_region	4317..4351 /note="REP01, repeated element" 6085..7200 /gene="ispb" /note="RC0724" 6085..7200 /gene="ispb" /codon_start=1 /transl_table=11 /product="octaprenyl-diphosphate synthase [EC:2.5.1.-]" /protein_id="AAI03262.1" /db_xref="GI:15619817" /translation="MNFVNHHCOKLKGAMOSGFGMRLLSQNPOEFLAMTEOPRIKIP MNIYIKIODDLKDEYTKLNNLITSLCKLDEBELIEVGRYLIEAGSKRIKRLPILITTA MVDYGDNHKIKLASAVEFTHTATLHDVVDSTLRKPIANVWGSRTSLIVGDFE ESQSRKPAVSGSIRKAMVNLAKASAIIISAGEVVOVLYKRNKQVNSKTSIIIVGDF ELFGTACAVAGAIIAEOVDHISAKMODFELKGLTIFEOVIDDLIDYDGNKQVNSKTS FLEGVATPLIFLFLYKRLAODKOLMTENIITCKDKPRKDPFVKIHDLMVKEIYNETVNY LSNEMASKLINKIPVQNIYKDYLFSTIKFILERSY" 6099..6158 /note="REP01, repeated element" 7207..7280 /gene="RCRNA17" 7207..7280 /gene="RCRNA17" /product="tRNA-Arg" complement(7405..7475) /gene="RCRNA18" complement(7405..7475) /gene="RCRNA18" /product="tRNA-Gln" complement(7527..9365) /gene="RC0725" complement(7527..9365) /gene="RC0725" /codon_start=1 /transl_table=11 /product="similarity to aminopeptidase" /protein_id="AAI03263.1" /db_xref="GI:15619818" /translation="WFAOPACKDRIIDIMTKIRINLLNLTETVEDIOGYIIPSDKRY SEYVSTAKRELYITGFTGSGNGIATITKOTLAFPTDGTLEQANKDELDFKLPDL DISKRGDAKIGYDELFETPTISLNKFNQKINGNLVDKILPMOPELNSKRYVLHDI KLVAGSHDKISKREIFLDSRFNCNDIEESGNEFSALVILDSSTICWMLNLRAS VAYLPMPAKVILITSTQLEIFLIPNTRDAEIIINAREPITLIPBEFENILNLSNTHT SKPARIRIIDTILASVHIMDLVADKVKQITDPCIMLCAKNVEIKIADILIKDAWA ALCEAFADFSCHPRENDGPEKHNNELFEYSGLKLTPEORAKOGVYSDSPATIGCG ENSAIIHYRADPKTKATTEGCGIILITSGGOCATADITPTIYIGKIPDOKRKYND VLKGHIALAKKAPKKNITAGANLILAOYILMOEMLDPICTGAGVGSFLSYHGGPOS ILNKRKTIILKAGMILISNPEGFVPGKGIQRIENMLYKENGWGLEFETLSLVPYASKI TDMTLNLINDEINIKYEYKNTAKIYDLSIQARWNLNNEINLYCN"
	gene	6085..7200 /note="ispb" /note="RC0724" 6085..7200 /gene="ispb" /codon_start=1 /transl_table=11 /product="octaprenyl-diphosphate synthase [EC:2.5.1.-]" /protein_id="AAI03262.1" /db_xref="GI:15619817" /translation="MNFVNHHCOKLKGAMOSGFGMRLLSQNPOEFLAMTEOPRIKIP MNIYIKIODDLKDEYTKLNNLITSLCKLDEBELIEVGRYLIEAGSKRIKRLPILITTA MVDYGDNHKIKLASAVEFTHTATLHDVVDSTLRKPIANVWGSRTSLIVGDFE ESQSRKPAVSGSIRKAMVNLAKASAIIISAGEVVOVLYKRNKQVNSKTSIIIVGDF ELFGTACAVAGAIIAEOVDHISAKMODFELKGLTIFEOVIDDLIDYDGNKQVNSKTS FLEGVATPLIFLFLYKRLAODKOLMTENIITCKDKPRKDPFVKIHDLMVKEIYNETVNY LSNEMASKLINKIPVQNIYKDYLFSTIKFILERSY" 6099..6158 /note="REP01, repeated element" 7207..7280 /gene="RCRNA17" 7207..7280 /gene="RCRNA17" /product="tRNA-Arg" complement(7405..7475) /gene="RCRNA18" complement(7405..7475) /gene="RCRNA18" /product="tRNA-Gln" complement(7527..9365) /gene="RC0725" complement(7527..9365) /gene="RC0725" /codon_start=1 /transl_table=11 /product="similarity to aminopeptidase" /protein_id="AAI03263.1" /db_xref="GI:15619818" /translation="WFAOPACKDRIIDIMTKIRINLLNLTETVEDIOGYIIPSDKRY SEYVSTAKRELYITGFTGSGNGIATITKOTLAFPTDGTLEQANKDELDFKLPDL DISKRGDAKIGYDELFETPTISLNKFNQKINGNLVDKILPMOPELNSKRYVLHDI KLVAGSHDKISKREIFLDSRFNCNDIEESGNEFSALVILDSSTICWMLNLRAS VAYLPMPAKVILITSTQLEIFLIPNTRDAEIIINAREPITLIPBEFENILNLSNTHT SKPARIRIIDTILASVHIMDLVADKVKQITDPCIMLCAKNVEIKIADILIKDAWA ALCEAFADFSCHPRENDGPEKHNNELFEYSGLKLTPEORAKOGVYSDSPATIGCG ENSAIIHYRADPKTKATTEGCGIILITSGGOCATADITPTIYIGKIPDOKRKYND VLKGHIALAKKAPKKNITAGANLILAOYILMOEMLDPICTGAGVGSFLSYHGGPOS ILNKRKTIILKAGMILISNPEGFVPGKGIQRIENMLYKENGWGLEFETLSLVPYASKI TDMTLNLINDEINIKYEYKNTAKIYDLSIQARWNLNNEINLYCN"
	CDS	6085..7200 /note="ispb" /note="RC0724" 6085..7200 /gene="ispb" /codon_start=1 /transl_table=11 /product="octaprenyl-diphosphate synthase [EC:2.5.1.-]" /protein_id="AAI03262.1" /db_xref="GI:15619817" /translation="MNFVNHHCOKLKGAMOSGFGMRLLSQNPOEFLAMTEOPRIKIP MNIYIKIODDLKDEYTKLNNLITSLCKLDEBELIEVGRYLIEAGSKRIKRLPILITTA MVDYGDNHKIKLASAVEFTHTATLHDVVDSTLRKPIANVWGSRTSLIVGDFE ESQSRKPAVSGSIRKAMVNLAKASAIIISAGEVVOVLYKRNKQVNSKTSIIIVGDF ELFGTACAVAGAIIAEOVDHISAKMODFELKGLTIFEOVIDDLIDYDGNKQVNSKTS FLEGVATPLIFLFLYKRLAODKOLMTENIITCKDKPRKDPFVKIHDLMVKEIYNETVNY LSNEMASKLINKIPVQNIYKDYLFSTIKFILERSY" 6099..6158 /note="REP01, repeated element" 7207..7280 /gene="RCRNA17" 7207..7280 /gene="RCRNA17" /product="tRNA-Arg" complement(7405..7475) /gene="RCRNA18" complement(7405..7475) /gene="RCRNA18" /product="tRNA-Gln" complement(7527..9365) /gene="RC0725" complement(7527..9365) /gene="RC0725" /codon_start=1 /transl_table=11 /product="similarity to aminopeptidase" /protein_id="AAI03263.1" /db_xref="GI:15619818" /translation="WFAOPACKDRIIDIMTKIRINLLNLTETVEDIOGYIIPSDKRY SEYVSTAKRELYITGFTGSGNGIATITKOTLAFPTDGTLEQANKDELDFKLPDL DISKRGDAKIGYDELFETPTISLNKFNQKINGNLVDKILPMOPELNSKRYVLHDI KLVAGSHDKISKREIFLDSRFNCNDIEESGNEFSALVILDSSTICWMLNLRAS VAYLPMPAKVILITSTQLEIFLIPNTRDAEIIINAREPITLIPBEFENILNLSNTHT SKPARIRIIDTILASVHIMDLVADKVKQITDPCIMLCAKNVEIKIADILIKDAWA ALCEAFADFSCHPRENDGPEKHNNELFEYSGLKLTPEORAKOGVYSDSPATIGCG ENSAIIHYRADPKTKATTEGCGIILITSGGOCATADITPTIYIGKIPDOKRKYND VLKGHIALAKKAPKKNITAGANLILAOYILMOEMLDPICTGAGVGSFLSYHGGPOS ILNKRKTIILKAGMILISNPEGFVPGKGIQRIENMLYKENGWGLEFETLSLVPYASKI TDMTLNLINDEINIKYEYKNTAKIYDLSIQARWNLNNEINLYCN"
	repeatL_region	4317..4351 /note="REP01, repeated element" 6085..7200 /gene="ispb" /note="RC0724" 6085..7200 /gene="ispb" /codon_start=1 /transl_table=11 /product="octaprenyl-diphosphate synthase [EC:2.5.1.-]" /protein_id="AAI03262.1" /db_xref="GI:15619817" /translation="MNFVNHHCOKLKGAMOSGFGMRLLSQNPOEFLAMTEOPRIKIP MNIYIKIODDLKDEYTKLNNLITSLCKLDEBELIEVGRYLIEAGSKRIKRLPILITTA MVDYGDNHKIKLASAVEFTHTATLHDVVDSTLRKPIANVWGSRTSLIVGDFE ESQSRKPAVSGSIRKAMVNLAKASAIIISAGEVVOVLYKRNKQVNSKTSIIIVGDF ELFGTACAVAGAIIAEOVDHISAKMODFELKGLTIFEOVIDDLIDYDGNKQVNSKTS FLEGVATPLIFLFLYKRLAODKOLMTENIITCKDKPRKDPFVKIHDLMVKEIYNETVNY LSNEMASKLINKIPVQNIYKDYLFSTIKFILERSY" 6099..6158 /note="REP01, repeated element" 7207..7280 /gene="RCRNA17" 7207..7280 /gene="RCRNA17" /product="tRNA-Arg" complement(7405..7475) /gene="RCRNA18" complement(7405..7475) /gene="RCRNA18" /product="tRNA-Gln" complement(7527..9365) /gene="RC0725" complement(7527..9365) /gene="RC0725" /codon_start=1 /transl_table=11 /product="similarity to aminopeptidase" /protein_id="AAI03263.1" /db_xref="GI:15619818" /translation="WFAOPACKDRIIDIMTKIRINLLNLTETVEDIOGYIIPSDKRY SEYVSTAKRELYITGFTGSGNGIATITKOTLAFPTDGTLEQANKDELDFKLPDL DISKRGDAKIGYDELFETPTISLNKFNQKINGNLVDKILPMOPELNSKRYVLHDI KLVAGSHDKISKREIFLDSRFNCNDIEESGNEFSALVILDSSTICWMLNLRAS VAYLPMPAKVILITSTQLEIFLIPNTRDAEIIINAREPITLIPBEFENILNLSNTHT SKPARIRIIDTILASVHIMDLVADKVKQITDPCIMLCAKNVEIKIADILIKDAWA ALCEAFADFSCHPRENDGPEKHNNELFEYSGLKLTPEORAKOGVYSDSPATIGCG ENSAIIHYRADPKTKATTEGCGIILITSGGOCATADITPTIYIGKIPDOKRKYND VLKGHIALAKKAPKKNITAGANLILAOYILMOEMLDPICTGAGVGSFLSYHGGPOS ILNKRKTIILKAGMILISNPEGFVPGKGIQRIENMLYKENGWGLEFETLSLVPYASKI TDMTLNLINDEINIKYEYKNTAKIYDLSIQARWNLNNEINLYCN"
	repeatL_region	4317..4351 /note="REP01, repeated element" 6085..7200 /gene="ispb" /note="RC0724" 6085..7200 /gene="ispb" /codon_start=1 /transl_table=11 /product="octaprenyl-diphosphate synthase [EC:

QY 79 GAAATTCGCGCTCTTCTTGCGCGATACATCTACGAGCTAAGAAAGTTCTGCCA 138
 Db 2796 AAAAGTTAAAGAAATATATGCGCTATAGAAAGAAATTAAGCTATTTTACTCA 2855
 QY 139 TGTTCCTAATGTTCTTCTGTATTTACATTTAACTATAGCGTGTTCGCGATACAAAGACA 198
 Db 2856 TGGCTTAAATGATGTATGATATCCCTTTTAAATTTGGGCTTTAGATCTATCAAGA- 2913
 QY 199 CTCCTATTTGCGAGCTCTGCTGTGCGAGAGCAATACCTTTCATCAAGTTTGGC 258
 Db 2914 ----TAGTTAGTACTACCTCTATAGCGGCGTGAATTAATGCTTTAAATTAATGCT 2969
 QY 259 TTGTTGTCCTCCGCTCTATTAATCTTATGCTTATGCTTATGCAAGCTAAGTATATTTAA 318
 Db 2970 TAGTATTAACCTCGTGGCTAATTTTACCGCTACTTTATGTTTAACTTATGATTAATCACTRA 3029
 QY 319 GTAAACAGGCTTATTTTATGAGTGGAGACGCCCTTTTAAATTTCTTGGCCGTTC 378
 Db 3030 ATTTTGAATATATTTTCTACATTAATAGTGGCAGTTTCTTACTATTTTCTTAAATTAATG 3089
 QY 379 CGACGTATTTATTCGCTAGCGCATGTTTACATCCACAGAAATTTGCTGACCGTTTAC 438
 Db 3090 CCTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3149
 QY 439 AGGCCATCTACCTCCAGGATTTGCTAGGACCTGTCATCTTAAAGAACTGACATTTG 498
 Db 3150 TAATTTGCTTACATACCTTATTTTAAATGCTTATTAATTAATGATGCTCAATGAGATTAATG 3209
 QY 499 CTGATTTTATGTAATCTCTGTAAGTATGAGGAGCGTCACTGCTATCTAATGTTCTGGG 558
 Db 3210 CACTGATGTAATTAATTTCCGGAATTAATGAGTACATGATCAATTAATGTTTGGC 3269
 QY 559 GATTTCCTAATGAATTTTCAAAAATCCACAGCAAGCAAGCGTTTCTACGCTCTTTTCGCTA 618
 Db 3270 AGTTTGAATCAATCAATTTTGTACTCTTAAAGCTAAAGATTAATCCCTTCTTGGAA 3329
 QY 619 TCGAGCTAATATTTCTTACTAGCTTGTGCTGCTGAATTTTGGGCTTCAAGTTGA 678
 Db 3330 TGGTGGTAATATTCGCTTATATAGCAGGAGCGTACTTCTTTTTCCTACACGGGC 3389
 QY 679 GAGCTTCCGTTTCTGAGGCTGATG------CTTGGGGAATTTCTT 720
 Db 3390 AGGAGCTATGATTCAGAAATTAATACCGATTTTAACTCATCTGCCGAATTCGTA 3449
 QY 721 TACGCTTTTGTATGCTATGATATTTATCTGACCTTGTCTTATGCGCACTTCTGCT 780
 Db 3450 TAATGCTTCAGCAATTCATGCTCAATTAATTTACTGCAAGAAATTAATGCTTCTAT 3509
 QY 781 GGATCAATTAAGAACTTATGACCGATCTCGCTTCTATATGCAAGAAATGCAAAAG 840
 Db 3510 TTAGAAATATAATGATTTATTTTAAACAGATTCATTAATTTTACGTGCAAAAAG 3569
 QY 841 GGAAGAAAGTGTCTAACCCTAATGAATGAAAGATAGCTTCTCTATCTTGTATAG 900
 Db 3570 TTACAGCTAAATGAAGAAACAACTTCGCTAATGAAAGTATCAATAGTAATTCCT 3629
 QY 901 CTCCTATATCTTTTATTAATCTCTGCTGTTATGCTTATGCTATTTGCTATTA 960
 Db 3630 CAAATATATAGCTGTATGCTTATTAATATCTGTTAGGATTAATTAATTAATTAATG 3689
 QY 961 TCGAAGTACTTGTGAAGATGCTAGTGAATGCAATATCTAATTTGAATGACTATAGT 1020
 Db 3690 TTGAAGAACTTGTGAAGCAAAATTAAGAAATTTACATCAACACTATAGCTATGTTA 3749
 QY 1021 AGTTATGAGGAACTTCTCTCTTGTGAGCTGCGTATACCTGCTATATCATGATTTG 1080
 Db 3750 ACTTATGAGGAGCTTATATTTGATGAGGATATCATGCTTACTTCAATGATAAAT- 3808
 QY 1081 TTGATGTAACCTCATTTGTAATTTTGTATGCTTAACTGAGCACTGATCATCTCTGCA 1140
 Db 3809 -GGAGTAATATTTTGTGAAGACCTGCTGCTATTTGCGCATTTATTAATTAATCTCTATTA 3866
 QY 1141 TGGTTCTCTTAACAGTATCGTTTCTGCTCTGTTATCTTTGAAGCAAGCTTCTG 1200

Db 3867 TGTTCCTATTAACGCGCTTGATGTTTATTTATCTTATATTTTATTTGAGAAATAGAG 3926
 QY 1201 GGTGCTGCTATGTTTGGTACACTCTCTCATCTAGCTGTGCTGTGAGGCTATAC 1260
 Db 3927 AATGTTGCTGATTTT-----AATCTCTTATATGAGCGCATTTATGTTGAGCAATTC 3980
 QY 1261 AGAATATCTTTGCAATCCAAATTAACGCTCTTTGACATCACTAAAGAAATGGCT 1320
 Db 3981 AGAATATCTGATTAATCATCAATATCTCTTATTTGATTTCAACAAAGAAATGGCT 4040
 QY 1321 ATATCCCTCTTACCAAGCAAGCAAAAGTAAAGCTCTATGATGATGTTGCG 1380
 Db 4041 ATATTCCTTTATCTTATAGAGCTTAGAAGTAAAGCAAGCTCCGATAGGTAATAGTA 4100
 QY 1381 CCGGCTTGGAAATCAGAGGAGCTTAAATCA-----ACAAAGTTTGTCTTATCT 1434
 Db 4101 CTAATATCGTAAATTCGTTGGAGCATTTATACAGCTTTGATTTATTTATTTCCGA 4160
 QY 1435 GTGAAGTATTTGAGCTATGACCCCTTATCTTGCAGTGTATCTTTTCAATGCTA 1494
 Db 4161 CTGCTACCTTTGATTTCTATTAATTAATTTATTAATACTTTTATGATTAATGATG 4220
 QY 1495 TTTGCTGCTTCTGCACTAAGTAAACAACTATCTTACGCGAGTCT 1544
 Db 4221 TATGATTTTGAAGCTTATTAATTAATGAATATGATGATGAGCTGTCT 4270

RESULT 33
 AE008626
 LOCUS
 DEFINITION
 Rickettsia conorii Malish 7, section 58 of 114 of the complete genome.
 AE008626 AE006914
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rickettsia conorii.
 Rickettsia conorii.
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia: spotted fever group.
 1 (sites)
 Ogata, H., Audic, S., Barbe, V., Arliguenave, F., Fournier, P.-E., Raoult, D. and Claverie, J.-M.
 Selfish DNA in protein-coding genes of Rickettsia
 Science 290 (5490), 347-350 (2000)
 2 (sites)
 Ogata, H., Audic, S. and Claverie, J.-M.
 Selfish DNA and the origin of genes
 Science 291 (5502), 252-253 (2001)
 3 (bases 1 to 10200)
 Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.
 Mechanisms of evolution in Rickettsia conorii and R. prowazekii
 Science 293 (5537), 2093-2098 (2001)
 4 (bases 1 to 10200)
 Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.
 Direct Submission
 Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France
 A public version of R. conorii genome database is accessible at <http://igs-server.cnr-mrs.fr/>. The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.
 location/Qualifiers

source 1. .10200
/organism="Rickettsia conorii"
/strain="Malish 7"
/db_xref="taxon:781"
73. .1608
/gene="tlc4"
/note="RC0666"
73. .1608
CDS
/gene="tlc4"
/codon_start=1
/transl_table=11
/product="ADP ATP carrier protein"
/protein_id="AAL03204.1"
/db_xref="GI:15619755"
/translation="MTINPSNVNSSKINSYFSKLDYTWPIKREHVSFLFTLLM
FCILFIONLRALKDSIVTWIGAEIISFLKFMGMPASFLMTAIVKLVKMAENI
FYLIISIFLFFELFAVYIPFNHEMLHSPVQNLMAISPLNKPIITLLSKSESLF
YIAELMNVFALLFWQFVNNTTVTESREYPLFGILISQGYIAGOPLENSIN
DYATNKRALOSSPHLSIQIILITVLVLLGIIATFPMILNKHVLDKHEMALFRKAK
KSMTASFOMLSSRHILATLULICYGAIINLVESGWKAATKIKTPEVAAFIG
SYLSYTGVEFIILEFVIGSNIVIRLGMFTAAVPLIVEITGILFFAVNNERFGLII
ANFLTDPALIAITIGAIQNVLSKSSYTFLEDSRTKEMAYVPLDEIKIKGAADVIG
TKLGKSSAFLOSLVFIILPSASYISTCLMIFITLCLTLMATKALNKEYNSIK
ESQ"
repeat_region 1691. .1724
/note="REP02, RS3-like repeat"
complement(1725. .1749)
repeat_region 1781. .1810
/note="REP02, RS3-like repeat"
1781. .1810
/note="REP02, RS3-like repeat"
1889. .4969
/gene="sca4"
/note="RC0667"
1889. .4969
gene
/gene="sca4"
/codon_start=1
/transl_table=11
/product="cell surface antigen"
/protein_id="AAL03205.1"
/db_xref="GI:15619756"
/translation="MRGFSKDGNDLTSEFDPLANKYEYEEOKOTLEBOKEFLSOT
TPALEADGFIIVASFAQSRTSPMSASGINSIPSDSPITKAVRETIITOPKDNLI
EOLIKDLAATITDRDLAOKREIEEKEKRTLISTFGNANREFIDKALENEBLKK
LESLIAGYKNVHNTSASGYPEGFKPVQENNVNSANDLRATVKNKADGDECTLWE
TVATKPTFLAKDGTQVQIISYREIDEPILKLDKADSGMLSMVALKADGTKCKDA
VYFAHYEBGNPKPOLKEISPKPLFAGCTGDALAVIHGGEIYVLAATRGYKEM
MKEVELNQGOSVLSQAEIIIGQSGSEOPPLITPOQTSSSVEPPQYKQOVPPIPT
NOPLOPESQPOSQOVNPLNLTATALSGMODLLNVYNAAGLTKAIDSKODLIRE
AATILNNEKSDIAEKQANITALEAVNNKNIKPKAKYVCVNAVULETTIKDONTPL
EKSKMEATYAIYLSNENLEPPKQOMLEKAVDGLSKDDASRAAIDIGIKDVIKIS
NLSPEDEMLIAGVDKVNSELSNAEKOKLGSYLKKGVEAQVSPAAQOIMQOHLVKI
MABOTKDTIKRVNDILFDELSTNELKTTNIOAITSNVLDGPATAEYKGEIIQAITWT
IAGSSLEAODKAAIINGVETIATHSPTSLSPNKALIMASAEKGIASEOTINPDEL
MTGGLVNGIIEGKGPEITKAVSSGIDNSNINSEKALKKADAASEALDPTDOL
TEGFKQNIIEHKPHDDIYNKAREVINAVPVIALEKSEKPEYSAEERTIQEOSTIL
NNTSKLAVERNPNRAWLSPNGLNKLIEEKKEEIKKDELVKAFTGKSTEEQOST
KTNLIDKTLSEKVRLOTIDKLQEOKRSEAIENPSVKTEDVRVSGSKSLKISKDN
PDIEKAKVVGDRVGRVKNIKIMGALMNARDIIQSENLNKSTPIKRESSPPQR"
gene
/gene="RC0668"
complement(5088. .5465)
CDS
/gene="RC0668"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03206.1"
/db_xref="GI:15619757"
/translation="MGRKNDIQKIDSFICKIYLSRLAKLSRQOLAEDIVDTHOOL
QKYEKAINRISVGLVLAELDRNIDYFPEGEEFANKPQVHTQHRMCIEVSNNEM
KINSTEEOAVNVLVKLACKKN"
gene
complement(5533. .5691)
CDS
complement(5533. .5691)

gene
/gene="RC0669"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03207.1"
/db_xref="GI:15619758"
/translation="MFFNNYLPNDYKRLKRCYFIYAVYNILMYFYLLDLKFNKILL
ITTCFII"
5987. .6814
/gene="RC0670"
5987. .6814
CDS
/gene="RC0670"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03208.1"
/db_xref="GI:15619759"
/translation="MKRIKNCVLALEFRTIEDGVEHSGYSEFILLISTPFLVLLA
LTSFLGASELQNFQIFLESLEPQATESIERKIRELLSAPQSLMNLAVISWTAS
SVECLRTIINRVYQIKSPPYIRLRLLSTIOPILSALITTFMFLVAPILFETKIP
IILEETIKRYIILNFIYRFLILIRLFGASSLYIILPNKLNFDVPGALLVYIIM
ISGLISTYIVYVYQNLMMYSGLSIIVTLIFFYIIMFIYGAEFNVLKNVENIE"
complement(6964. .7797)
gene
/gene="RC0671"
complement(6964. .7797)
CDS
/gene="RC0671"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03209.1"
/db_xref="GI:15619760"
/translation="MKISHINYSFINTLSGRLOKNKOLEKIIIAOFHODYNIPIIGV
EIEFYLSHNIDIKFELISRRYLARFIKIKKEKGNQOEIDPVSNILOYIKNL
EVKTIKKMAQOLNGYIDPSPKPLDYGSMHPHINSEFNNDYIILAAOGLCHYM
LDPLIAPMTITDYARINKKEMAPTHISYCGNRSVAVRPNAPKRLERLSSPND
PYIAITTIKSIILALKSPSSLOTIKETIGNADPOYNLPLPSSQASFMILKRPDF
K"
gene
/gene="RC0672"
7796. .8518
CDS
/gene="RC0672"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03210.1"
/db_xref="GI:15619761"
/translation="MEGLINSGYVYKLFDEKFNNSHRYIKNNSTINEPIVENKSI
TYFEKODIILNVOHGNQIVNADESIIVPEADGSTTKNLIIVAGSALDVCYVILA
SGGKITGAHACHKGSINNIISNIYTKITEKAKNLIIVGALIAOSSIEYVDEYK
APLSKIDNNKQFTHSIKENHYMDPAFVELKLEKXGVKDIRINADITYTNLPKYP
KRSYHLOEYVNONILSAIVMK"
8597. .8630
/note="REP01, repeated element"
8622. .8786
/gene="RC0673"
8622. .8786
CDS
/gene="RC0673"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL03211.1"
/db_xref="GI:15619762"
/translation="MAIQIKIKFKCLEFTTGILLHQLLCNFLVMTGKPYATMLARNN
DIVLYLAMOK"
8650. .8687
/note="REP01, repeated element"
8811. .9362
/gene="def2"
/note="RC0674"
8811. .9362
gene
/gene="def2"
/codon_start=1
/transl_table=11
repeat_region
repeat_region
gene
CDS


```

/product="polypeptide deformylase [EC:3.5.1.31]"
/protein_id="AAL03212.1"
/db_xref="GI:15619763"
/translation="MNQDKPYQYLYVAANDIFFKQAEYIDIVDDNIRIYDKMLQNLH
IERAGDGLANWGLIKRTIAVVDIHENMNSSPIVLPNPIYFSEKQFTFESLSPFG
IEASTTRSKAIKVKLYDINGNKQELAAEGFLATVYQHIEIYLNKRTFLDLSKIKRDT
LKKMLKIKHKLHPPIHGGGRH"
/complement(9295..9522)
/gene="RC0675"
/complement(9295..9522)
/gene="RC0675"

```

Query Match	13.2%;	Score 215.8;	DB 1;	Length 10200;
Best Local Similarity	49.7%;	Pred. No. 5e-41;		
Matches 668; Conservative	0;	Mismatches 662;	Indels 15;	Gaps 4;

OY	74	TTTGGGAAATTTGGGCTCTTTCTTGTGGCCGGATCATCTGACGAGCTAAAGAAAGTCT	133
OY	126	TTTTTCTTAACCTTACCGCATTAATCTGGCTCTTAACGCCATGAGATTCTTAATTTTT	185
OY	134	GCCATGTTCCTAATGTCTCTCTGATTATACATTTAACTATACGAGTTACCGGATACAA	193
Db	186	ATTCATCATTATTAATGTCGTATTTTATTTATCCAAATCTAATCAGAGCTTTAA	245
OY	194	AGACACTTTATTTGGGAGCTCCTGGTTCTGTGTGCAAGGCAATACCTTTCATCAATT	253
Db	246	AGATAGATTTGTTA-----CTACTATGATCGGTGCCAGATTAATCTGTTTTGAAAGT	299
OY	254	TTGGCTTTTTCGCCCTGCTGATATCTTTATGCTTTATTTATGCAAAAGCTAAGTAAT	313
Db	300	TTGGGAGATGATGCCGTAGCTTTTTTATATGACTCTATATATGTAAAGCTTGTATATA	359
OY	314	TTTAAGTAAGCAGGCTTAATTTTATGACGTGGGAAGCCCTTTTAATTTCTTGGCCT	373
Db	360	GATGAAGACAGAAAATATATATTTATCTTATATATCAATATATTTTAACATCTTGGCT	419
OY	374	GTTCGCCACTAATTTATTCGCTACGGGATCTTTTACATCCTACAGATTTGCTGACCG	433
Db	420	TTTTGCTTAAGTATTTTTTCCAAATCATGAATAATGCTGCATTTCCAGTCTCTGTAACGGTTA	479
OY	434	TTTACAGGCCATCTACCTCCAGAGATGTCGAGACCTGTGGCATCTTAAGAAGCTGAC	493
Db	480	AAATTTAATGGCAAGTTTACCATAATTTAAATGTTTTATATGCGCTTTTATCAAAATGGAG	539
OY	494	ATTGCTGCTAATTTATGACTTGTCTGAACATATGGGAAAGCTCATGCTATCTTAATGTT	553
Db	540	TTTTTCGTTATTTATATAATAATAGCCGAATTTATGGCCAAACATGATTTTCGATCTCTTT	599
OY	554	CTGGGAGTTTCTGTAATGAAATTAACAAAATCCAGCAAGAACGCTTCTACGCTCTTT	613
Db	600	TTTGGCAGTTTGTAAATAATATATTACTACAGTAGAAGAAATCTAAAGATTTTATCCGTTAT	659
OY	614	CGGTATCGAGACTAATATTTCTTACTAGCTTCTGTGTGTCGAATTTGTTGGGCTTCANA	673
Db	660	CGGTTTACTTATGTCNAACAGGATATTATTTTACAGGGCAGTTTGAANAATCTAAGTAA	719
OY	674	GTTGAGACTTCCTTTCTGGAAGGTGTAGATCCTTGGGGAATTTCTTATAGCTTTTGAT	733
Db	720	TATTAATATATTAACCTCAATAA--TAAATTTGCAATTCGAATGCTTTTTCATACACTTTC	776
OY	734	GGCATGACTATTTGATCTGCACTGTTCTATGCGCCAGTACTGCTGATCAATTAAGAA	793
Db	777	TATACAAATTTATACAACTATATAGTTAATTTTAGMAATATATAGCTATTTAAACITTTTG	836
OY	794	CGTATTCACGATCCTCGCTCTTATATTCACAGAGAATATGCAAAAGGGGAAAAAGGTGC	853
Db	837	GTTGCTTATATCAATAAGTACAGACAAAGAGACATTTGGCTTACCTCAGGTTTAAAGCAA	896
OY	854	TAAACCTAAATGATATGAAGATATAGCTCTCTATCTGATAGATCCTCTTATATCT	913
Db	897	GAAAAAATCTATAGCTATTTGCGGAAGTTTTAGATGCTTCTATGCTCAGACATATTTAG	956
OY	914	TTTATTAACCTCTTGGTTATTTGCTATGATATTTGCAATTAATCAATTCGAAGTGCATTG	973

Db	957	ATTATATGGCAACCTTGGCTTATCTGCTATGCGCATGGCCATGCTAATTTAGTAGAAGGCTCTTG	1016
Qy	974	GAAGAAGTCAGCTGMAAATCGMAATATTCCTATATATGATGATCTATPAGTAGCTCATGAGGAA	1033
Db	1017	GAAGAAGCAGCAGCTACTAAATTTATATMAAATCCAAACCGAATATGACGCTTTATATAGAA	1076
Qy	1034	CTTCTCCTCTTGAGACAGCGGCTAGTATCCGTAATCATGCTATTTGTTGGTGTAAAGT	1093
Db	1077	TTATTTGAGGTATAC---CGAGTATTTACTATTTTATTTGTGCTACTTGGTTCAAATAT	1133
Qy	1094	CATTCGMAAATTTGSGATGTTAACTGGAGCCCGACAGCACTCCTGTCATGCTTCCCTAAC	1153
Db	1134	AGTTAGAGACTTGGCTGGTTGATTACGGCGGCTGTATACCCCTTTAATAGATTTTATATAC	1193
Qy	1154	AGGTATCGTTTCTTCGCTCTCTTTATCTTTAGAAACCAAGCTTCTGGGCTGGTGGCTAT	1213
Db	1194	CGGTATATTTATTTTTCGCCGTTAATTTTGAAGAATTTGGCGGCTTAATATATAGCAA	1253
Qy	1214	GTTCCGCTACAC--TCCTCTCATGCTAGCTGGTGGTTCGGAGCTATACAGATATCTT	1270
Db	1254	TTTATTCTTCAACCGATCTCTCTTAACTTAACTAATCAATAGGTGCTATTTCAAAATGCTCT	1313
Qy	1271	TTGCAAAATCCAAATATAGCTCTCTCTTTCAGCTAACATAAGAAATAGGCGTATACCCCT	1330
Db	1314	TAGTAATCAAGCAAAATATACCTTATTTGTTTCACAAAGAAATGGCTATGTTCTTT	1373
Qy	1331	TGACCAAGAGCAAAAGCTCAAAAGTAAAGCTGCTATTATGTAGTGTGCGCCGCTGTGG	1390
Db	1374	AGATCCGGAATATAAATTAATAAAGCAAGGTGCTGCCAGCTGATAGGTACAAAACTCGG	1433
Qy	1391	AAAATCAGAGGAGGCTTAATCCA	1415
Db	1434	TAAATCTGGTAGTGCAATTTTAA	1458

RESULT	34
REFX02/c	
LOCUS	RPXX02
DEFINITION	Rickettsia prowazekii strain Madrid E, complete genome; segment 2/4.
ACCESSION	AJ235271 AJ235269
VERSION	AJ235271.1 GI:3868717
KEYWORDS	Complete genome.
SOURCE	Rickettsia prowazekii.
ORGANISM	Rickettsia prowazekii Bacteria; Proteobacteria; alpha subdivision; Rickettiales; Rickettsiaceae; Rickettsiidae; Rickettsia; typhus group. 1 (bases 1 to 312430) Anderson, S.G., Zomorodi-pour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G. The genome sequence of Rickettsia prowazekii and the origin of mitochondria Nature 396 (6707), 133-140 (1998)
TITLE	
JOURNAL	Nature 396 (6707), 133-140 (1998)
MEDLINE	99039499
PUBMED	9823893
REFERENCE	2 (bases 1 to 312430)
AUTHORS	Anderson, S.G.E.
TITLE	Direct Submission
JOURNAL	Submitted (11-Nov-1998) S.G.E. Andersson, Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN On Nov 13, 1998 this sequence version replaced gi:3860788.
COMMENT	
FEATURES	
source	Location/Qualifiers 1..312430 /organism="Rickettsia prowazekii" /strain="Madrid E" /db_xref="taxon:782" 76..1830 /gene="RP226" 76..1830 /gene="RP226"
gene	
CDS	

/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14689.1"
/db_xref="GI:3860789"
/translation="MNLKCDILFLINLLVTVGYASPPPLPSPSLPIIADVTDKNI
STNSISFEKQFESKQKKKNISQEQEQOTKAIHQESQOIDSREINENQSEPTI
DFGSTITPVSASNYIDSKAEXENSTNLAASYNTQDIOYKQDFPSEASBIDIGNTK
FTSATNHEMKYEAASNDKENTLNTSNTITPWPSPVLSIPTADVNYYVPSQSOVQIY
KPTNLTSIRNPILPNHHTDLNKEKNELESTISNMTTIPTNWSVPSIDDTIOTLTNIT
VPTAEHTSVQTSYVMHSNOHSAOPIPTISINTPVETSSVTLRATESVPTVINSQEIF
VSESESTRKQDWTPIPMVLYVDPKSSQSLALQKNNQDITINNOAESVSSSNV
TIQKONQVNNATSESTREKYNETOMLPEDDDIVGKLTEDATLEQMDHGYTKLF
QKREEMIANAEKRLVSEFIKYDDINKNDIYANLSYSAVDNAFRAVENNNFEELR
ALLDVPYLIQAKNSTGETLTLTSIYNGNYLAKELVIRGITSVLKNDCEKYPIDIAL
AKGNTNYCMKMKAKGYN"
2025. .4013
/gene="Rp227"
2025. .4013
/gene="Rp227"
/codon_start=1
/transl_table=11
/product="DNA GYRASE SUBUNIT B (gyrB)"
/protein_id="CA14690.1"
/db_xref="GI:3860790"
/translation="MSDLFSLNKKKKNIYVYNSANDIEVLGLPEYRKRPGMYIGT
TDSNAMHVLSEVLNDAMDEAVAGFASIIIMIKMHDSITTFDNGRGIPIDNHRFPD
KSALEVLITLTHSGSKFSNNYHTSGGLHGVGISVYNAASKHFIKYKQKGLYSQSY
SKGAKLIDLISAESKRLRGTSINFTPDPEIFESKRLHPNKIETIASRKAIVLGSYS
IEMCEVEYPSDIPKKALINPNPKLDKLSKISLDNLVPEIFSGNESTVDDIKLE
WATCQNNNDISAFMOSYCNVPTPQGGHEGSLKALIRGLKATSEMIGNKKAANLTI
EDILETASIVLSIFIVEPSFOQTREKIVSGVSKLVENIIRKDHFLSDKVLATH
LLEHVIAIAEERRNKKERNISRSKSVTQKLRPLGKADCTRTSAEGTELFVEGDSAG
GSAKQNRRETQAVPLMGVLANSTLEKTIINNOAIODEIALACGSKLKNYKKNL
RYEKTIIMTDADVDAVGHASLMTPEFLRMPKLYVEGHLVAKRPLRYLRQSNKIYVA
CDEEEKIKITKYLSKASKAKTEVGRFGLGSMAPQLEKTTMHPKRSLLKVTLEDVO
NWDKIVDDLMGSKPEKRPQFIYEOALVAMDITINKLDI"
4122. .5447
/gene="Rp228"
4122. .5447
/gene="Rp228"
/codon_start=1
/transl_table=11
/product="Tail-SPECIFIC PROTEASE PRECURSOR (ctp)"
/protein_id="CA14691.1"
/db_xref="GI:3860791"
/translation="MLLRFITALFLSINCAIEGKETENKISNOEAYKOPQDVEIRIK
DYVQVDPKQIMDAINGMNLSDPHSNYVTDDELDTFTPTKGEFGIGVEINYDSG
AIKTISSIDLPAPKAGIKGGDIYGVNDELVSTLGPVKATKEGRGTPGKVRLLTIK
EEBAKPOLELTREIVIKIPYKHALEKNNTAVIRITTFNESTISELKAAYKRLTESK
DNLGIIILDLNNGGILDOAIVASDYFIDSGVITFGRTSSNSETKANFEFLKAP
KVPWIVLINGSASASEIVAGALODKRAIILGTSGFGSGVQALTOINSAARKLTI
SKYVTPSGRSIOAGIEPDLIPPAKVEYEPVKIDKRFSSSLKNVLIKNDMANKNOS
NKEETKNNKQSESELSEIYKKQYAFARADVITGLINTNLETOGAK"
5653. .7146
/gene="Rp229"
5653. .7146
/gene="Rp229"
/codon_start=1
/transl_table=11
/product="HISTIDINE KINASE SENSOR PROTEIN (bara)"
/protein_id="CA14692.1"
/db_xref="GI:3860792"
/translation="MNHNKYIVRLSFTILFLNLYNMLEFYEMIKEMIKOVALERT
KIYELYDINIMNHIKRVISKLHKFDYDLLODDQFINFVKITAQMPNLNINISLYLT
KGNKEITSNMLHMSYDNYKDDSLIEIVTKIKDILKSFESAAPLADAKGITSHTL
LPKVIENESDLTEHAHSVTSYIPIVDHNDLPYDAVFETINTNITSQNNITSLEAK
VFTEFIITFIETFTIISNTNVAQOIIIEQDLETNRNLKAOIEKERTSSNTREFAN
SHELRTPLNALIIGSEILMSERDEPKSNYIKDIDHGKHLISMINDILDSAAKADK
LKVDNIDLNLKLISSELILIKPRAYQAEVLLSRLEKHEHVINADPRKROVYLLNL
SNAYKTPSGSVTISLEKDELAKLYIKVINDYIGIEEKDIPKTLISAFQIDSELSR
KYEGTGIGLPLTKVLVLMNGKPDLOSKINKGTITVTITTKYDGSIEL"
7374. .8126
gene

CDS
/gene="Rp230"
7374. .8126
/gene="Rp230"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14693.1"
/db_xref="GI:3860793"
/translation="MKLITLFLSFLSLSGSKSEMPQGRFLKAYAVNNDFENRLDEQ
QEIRLIGVETLIDHKITLIDKONLMNLOEENITVELTDEPKKODITFDIALBEGMHN
VSKRTPEFNKRIAYKQAYDALAYKDNKITEAKDKFNQIQYPPNNSLISNAYFY
AEQFKQKDNIGALINLYKCYKESPKAKSSDGLLKALSLISGELKKQAEACNIIAKLD
KEPFINRTSVSKMTEDAKIKFGCKINKNNKI"
8123. .8785
/gene="Rp231"
8123. .8785
/gene="Rp231"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14694.1"
/db_xref="GI:3860794"
/translation="MTDILDEVLSQDNEKRLIFKFLVPIIITISIVITIMYITIN
NKDNRIKNNQKNDILVKTIGLDTTKONKALAEFTLENLTSSMTKIREKTALEQVAI
RMSERKYLGAKNLNLNKIIDNEKSEISTAVARIAMCIVLDDQSLNIDQKGLQKYLIN
YFDDERKPFMTATITIKAIIDIKHNMTOAEKNNKLLASNNVSDLLKQAKALLVSL
SK"
8812. .10059
/gene="Rp232"
8812. .10059
/gene="Rp232"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14695.1"
/db_xref="GI:3860795"
/translation="MKRKLALLLPILISGNGIGSKSVKNIVDLTPKLYIQTNEPY
LDSNTKIPFNVMNLKKNQYSPASKMIAEVPFIDGMIYTLDIRANISAFSIEKNKI
WSYVLSKHKNDNYIGGILHNGKLYTYAGRLIYVLDASGEIIRKELPDIRIRP
IALNDHTLIQVOTISNOTIALDSELTAKTLMHESIATLSTSYSMIPVQHDNVYVYN
TGOVLATINNGEYKKNPEFTNLNDHRAINFDPSISLTCPYHDSMLYIATGCGIKI
KNLITGSVMQINADDIQSSLSIGNSLFIINNAQIALNPETGAKFYADVLYEIN
DKRLSTPFLVPEYGVNNOGRLNIVSDGILYNNFIDNSGLKMPHIIKIKINRY
YGLRSNNTLYFSTDROYIFGSOI"
10279. .10746
/gene="Rp233"
10279. .10746
/gene="Rp233"
/codon_start=1
/transl_table=11
/product="50S RIBOSOMAL PROTEIN L13 (rplM)"
/protein_id="CA14696.1"
/db_xref="GI:3860796"
/translation="MKTIYSAKPSIEKKMYVIDAKNVYLCRLASRAVIMLRGHHKPSF
TPHDCGDNITIIINAEHIKTLGKLNPKDKVYRHTGPFGGIKDTTAAGITLGGKYPE
RVIKMAVKRMITRNVLGAKOMSNLYVANCHEHPMAQOPIYDFASENPKNK"
10753. .11238
/gene="Rp234"
10753. .11238
/gene="Rp234"
/codon_start=1
/transl_table=11
/product="30S RIBOSOMAL PROTEIN S9 (rpsI)"
/protein_id="CA14697.1"
/db_xref="GI:3860797"
/translation="MTELKIKTEKVVAKQLTKRESIKSVYKLPKIKIDSYSKFFATYGRK
NATARWLVKVGKRIYVNNKILNQVPSSEYVVTIIQPIITLTFTDQYDVCIVKGG
ISOQKAILHGISIKALDKSAPCEHAIRLGLLTPDSRVERKKYQGRKARKKTQESK
R"
11279. .11352
/gene="tRNA Met (CAT)"
11279. .11352
/gene="tRNA Met (CAT)"
gene

```

gene      complement(12284, .12769)
          /gene="RP236"
CDS        complement(12284, .12769)
          /gene="RP236"
          /codon_start=1
          /transl_table=11
          /product="INVASION PROTEIN A (Inva)"
          /protein_id="CA14698.1"
          /db_xref="GI:3860798"
          /translation="MRSSNKYLDLPYRGVGMILNADNOIFVKRIDTKISSWOMP
          OGGIYVGETPSIAAREMLEEIGSNKGIIEISCMYSDPSLIPKLMNGNFRGOK
          ORWFLIRGTGNKKNINHTSNPEPDQMWISLDELSTIIFPKRLXYAVVKEFESLI
          Q"
gene      complement(12798, .14150)
          /gene="RP237"
          /complement(12798, .14150)
          /gene="RP237"
          /codon_start=1
          /transl_table=11
CDS
Query Match      13.1%; Score 214; DB 1; Length 312430;
Best Local Similarity 48.4%; Pred. No. 9 9e-41;
Matches 720; Conservative 0; Mismatches 755; Indels 13; Gaps 4;

```

```

Db 181571 TCTTTAAATCATTCATCAGATTAATTTATATCAGATTAATTTCTTAAAGCTCTCC 181512
QY      787 ATAAGACGATATTGACCGCATCTCGCTTCTAATATCCAGACAAATGCAAAAGGGGAAA 846
Db 181511 ATTAGCTAATTTGATTAATCAGTTGTAGAACGTGATATAAATTTAAATCAACAAA 181452
QY      847 AAGGTCTTAACCTAATAATGAATGAAATAGATGCTCTCATCTGTGATACATCCCTT 906
Db 181451 GACACATATATTTAAATTAAGCTTGCTGAAAGTGCAAAAATTAATCTTAAGCTAGAT 181392
QY      907 ATATCTTTTAAATTAACCTCTTGATATTGCCATAGATTTGCAATTAATTAATGAAAG 966
Db 181391 ACTTGCTTTATTTGTCTCTTCGTAATGTCTATTCTATGAGTATTAACCTAATAGAAG 181332
QY      967 TCACCTTGAAAGTCAAGCTCAAGCTCAATATCCTAATATGATGACATATAGTACTCA 1026
Db 181331 GATTGGATGTCCAAAGTAAACAACTCATCTACGCTCAAGAGATTTATATCATATC 181272
QY      1027 TGGGCACTTCTCTCTGACTGGGCTGATACGCTACTATTCATATGATTTGTTGCTG 1086
Db 181271 ACGGTGAAGTATTTGTTGGACTGGAGTGTACTTACTTATAGTG--CATTTTATAGCA 181215
QY      1087 GTACGCTCAATTCGTAATTTGGATGTAACTGAGGCCCTAGTCACTCTGTCATGCTTC 1146
Db 181214 GTAGTTTATTTAGAAATTTATGCTGTTTGGGGGCTATTATACACCGATATATAGCT 181155
QY      1147 TCCTTAACAGTATCGTTTCTTGCTCTGTTATCTTTAGAAACA--AGCTTCGGGC 1203
Db 181154 TTGTAGCAGGGGTATATGTTTTTTCATTCACATTTTTCACAACTAGGAAATATAG 181095
QY      1204 TGGTGCATATGTTGGTACCACTCCATCATCTAGCTGCTGCTGGAGCTATACAGA 1263
Db 181094 TAAATATCTTGTGGCTATATCTTCCACTTGCTCATTAATATTTTATGGTGACCTTGGC 181035
QY      1264 ATATCTTTTGAATCCAAATATACGCTCTCTTTGACTCACTAAAGAAATAGGCTATA 1323
Db 181034 ATGTATTTGCTAAATCTGTAAGATATTCCTCTTTGATGCTCTAAAGAAATAGTGATTA 180975
QY      1324 TCCCTCTTGACCAAGACAAAGTCAAGGATGAGCTGCTATATGATGTTGCCGCC 1383
Db 180974 TTCCACTAGATATGAAATTAAGACTAAAGGTAAAGCAGACTGATATAGGGTGCTA 180915
QY      1384 GCTTGGAATATCAGAGAGGCTTAATCCACAAAGGTTGCTGATATCTGTGGAAGA 1443
Db 180914 AATTTGCTAGTCAATAGTGTCTATTAATCAATTAATATCCTTATGATATCTTCCAAATG 180855
QY      1444 TTGAGCTATGACCCCTTATCTTGACATGATTCCTTTTCATCATTTG 1491
Db 180854 CTGTACATATACGACATAGCAGCTTATGATGTTACTTTATTTATGCG 180807

RESULT 35
AE008614/c 13627 bp DNA linear BCT 14-SEP-2001
LOCUS      Rickettsia conorii Malish 7, section 46 of 114 of the complete
DEFINITION genome.
ACCESSION AE008614 AE006914
VERSION    AE008614.1 GI:15619590
KEYWORDS   Rickettsia conorii.
SOURCE     Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;
ORGANISM   Rickettsiaceae; Rickettsiaceae; Rickettsia; spotted fever group.
REFERENCE  1 (sites)
AUTHORS    Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E.,
TITLE      Selfish DNA in protein-coding genes of Rickettsia
JOURNLM    MEDLINE 20485642
PUBMED     11030655
REFERENCE  2 (sites)
AUTHORS    Ogata,H., Audic,S. and Claverie,J.-M.

```


gene	LSLDSAKVILTSYELGFIOLCLNWSYMSLSLEGLMMSRYKOLYPTAPKOFIAHGKAV PEWGLILVSAFLGSLIICLWMSYMSLSLEGLMMSRYKOLYPTAPKOFIAHGKAV LGSAPLUVIVFIFGGLMHWLSKSVKSLPATEKMTIPLDSEKTKTGAAVDMGAK IGSISGLAIQFISFISIPNNAIHDIAGLMFSFIICLLMLYGVKSLKYNNKIOR" complement (6347 . 7960) /gene="pyrg" /note="RC0523" complement (6347 . 7960)
CDS	/gene="pyrg" /codon_start=1 /transl_table=1 /product="ctp synthase [EC:6.3.4.2]" /protein_id="AAL03061.1" /db_xref="GI:15619600"
gene	/translation="MVFHIFVTGGVSSLGKGLTASLAMLLOAKGFRVSRKIDPYL NIDPGNMPHEHGEVYVDGAEFTLDLGHYERTGYSAKGFSDITGAYLSKLDDE RLNGAVTQVIVPHTNIIKDFLISDTGDFICIGIGFVGIELPFEALROIQIG NKLSENCLEFHEFLRLPYKTARELKIPROHSYKALRAIGSPNIIICVAERISG ELIKSLICNIESPVYVPAIDOKNIVYPLAIVHNSGDMVKLEFNNIMPSDKRHH DIIRLDSNSKVAIATIAKTHKLDKQKSYIELDNAGIYIKKIDLVMANLIE ENIRKLKLIDDEGLIVPGFEFERATKGTIIAKYARTNNIPPEGCFEMOLATIEI LIGKDAVTEEGFKVDKGTIIIEKINKNEDESKITENYKTMRLGSLGSCSVASIA AYKSLIEFHEHRIHRYKFNNEFQNIIEFKHGIVTSEGSDEIETELIELPLRMFVG HPEKSKPEFHHPLPIFOIKAAIENYKCN" complement (7963 . 8703)
CDS	/gene="kdsB" /note="RC0524" complement (7963 . 8703) /gene="kdsB" /codon_start=1 /transl_table=1 /product="3-deoxy-manno-oculosonate cytidyllyltransferase [EC:2.7.7.38]" /protein_id="AAL03062.1" /db_xref="GI:15619601"
Query Match	12.9% Score 211.8; DB 1; Length 13627; Best Local Similarity 50.5% P. Ident. No. 4,4e+40; Matches 722; Conservative 0; Mismatches 672; Indels 37; Gaps 7;
OY	1 GAAATAAAAAAGCTATGAGATGAAAAATGATTTTCAGAGGT-AAATATGACAAA 59
Db	6165 GTACTAAAAAAATTAAGCTCATATGTAATCAATGCTAGCCGGATTCATTAATTTCA 6106
OY	60 ACCGAGAAAAACCTTTTGGAAAAATTTGGCGCTCTTCTGTGGCCGATCATATCAG 119
Db	6105 TGGAAATTAAGCGACAGAAACAGTAATATTGAGCATATAGATGGCCGATTAAGATCGATGA 6046
OY	120 CTAAGAAAGTTCTGCCATGTTCTCTAATGTTCTTCTGATTAACATTTAACTA 179
Db	6045 TTAACCAAGTTCAATCCCATGGCTTTACTAATGTTTTTAATCTTAATCAAAACTTA 5986
OY	180 TTACGCATACAAAAGACACTCTTATTTGTGGAGCTCTGCTGTGGTGGACAGGCA 239
Db	5965 GTTCGTAGTATTAAGATGATGTTGTGTG-----TACCTTAATTAAGCTCGA 5932
OY	240 CTTTCATCAAGTTTGTGCTGTGTTGCTCCCGTGGCTGATTAATCTTATGTTAT 299
Db	5931 AGTTTATCAACACTTTGGGGGAATGCGCATGGGAATTTATTTGTTATTTTATCT 5872
OY	300 AAGCTAAGTATTTTAAGTAGACAGCCCTTATTTTATGCACTGGGAACGCCCTTT 359
Db	5871 AAACCTCTGATATATTATGACAAACAGACAGTTTATAGATAATTAACCGGTA 5812
OY	360 ATTTTCCTTGCCGTGTCGCCGACGTATTTATACCGTACAGGATGTTTTCATCT 419
Db	5811 TTTTTCCTTGCGCAATTTTGTGTTTATTTTATTTCCGTACAGAGATTTTTCATCCGAT 5752
OY	420 GAATTTGCTGACCGTTTACAGGCACTCCCTACCTCAGAGATTCCTAGACTCGTTGCCAT 479
Db	5751 CCCGAATTAATTAAGCACTATATACGCTGTCTCTCACTTAATAATGGTTTTTAATATT 5692
OY	480 TTAAAGAACTGGACATTTTCTGCATTTTATATGTAATGTAATVGGGAACGCTCATG 539

[illegible]

Db 23321 GTACAAACTCGGTAAATCCGGTACTGATTTTACAA 23384

```
RESULT 37
RPPX04      237523 bp   DNA      linear   BCT 11-NOV-1998
LOCUS      Rickettsia prowazekii strain Madrid E, complete genome; segment
DEFINITION 4/4.
            AJ235273 AJ235269
            AJ235273.1 GI:3861237
            complete genome.
KEYWORDS   Rickettsia prowazekii.
            Rickettsia prowazekii.
            Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.
SOURCE      Andersson, S.G., Zomorodipour, A., Andersson, J.O.,
            Sichteritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K.,
            Eriksson, A.S., Winkler, H.H. and Kurland, C.G.
            The genome sequence of Rickettsia prowazekii and the origin of
            mitochondria
            Nature 396 (6707), 133-140 (1998)
JOURNAL    99039499
MEDLINE    9823893
REFERENCE  2 (bases 1 to 237523)
            Andersson, S.G.E.
            Direct Submission
            Submitted (11-NOV-1998) S.G.E. Andersson,
            Srv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University
            of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
FEATURES
Source      1. .237523
            /organism="Rickettsia prowazekii"
            /strain="Madrid E"
            /db_xref="GI:3861237"
            /protein_id="CAAI5137.1"
            /db_xref="GI:3861238"
            /translation="MAEPIIFTLFGKDRILDRINNOKALNKLHIIPLOVNESHS
            DKERIEILEKSYKECKIPFIENIOISNNVKKPIFRAODLIINAONLKITTEDEKN
            ALPONSLEKSYKMOIIEKVDHVFETNADONLSIADGIVPKDKTTTIDTIDRLTSVN
            NLIYNHETNHGSLSTIYKKVDKLITAKKODSKELIETPISVDEATNLTLAGTIS
            EDOYIGLKEIHELKDPNNAENFKYVQISQKFKDKDKOKKHIDSNFNTQKVI
            NDSNKDIOVEHITSYVODKESKPOGFLTRIFENYKNIYVKO"
            2074. .3348
            /gene="Rp702"
            2074. .3348
            /gene="Rp702"
            /codon_start=1
            /transl_table=1
            /product="unknown"
            /protein_id="CAAI5138.1"
            /db_xref="GI:3861239"
            /translation="MPTLIIVIIIMISLSALFAETATATASSPKIHKLITAGNKR
            AKTVLEVLKKKEVIGILLIGNSLINTCTTATLTLSLGNNGTIVAGVAFIT
            VRAEYVKATAIVAKPEDLTKMSSTIVFLKLRPIINADYITKJECFIRINLNO
            ISGTEEVGVIEYHOGGVYKSERNNVGLIIRNMTVSEIMTHRSNIIALNIDLPY
            ELIKITLISGAHRIPLMDNRNDIIGILNKLKALYENNDKTKVDINTLTPEPM
            FTIDNALVVDQLHAERFRRNNHAGVDEYGLTGLITLEDVIEIYVAGTIDDEHRLNN
            EIIQESMTERTIKTITRIDINRELDMNLSDEPANTAGLIHKIARIIPNOGEVITKE
            NFKIILKRIANKIDSVKITVLPMTERTISSE"
            3341. .3353
            /gene="Rp703"
            3341. .3353
            /gene="Rp703"
            /codon_start=1
            /transl_table=1
            /transl_table=1
```

```
/product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
(ccmf)"
/protein_id="CAAI5139.1"
/db_xref="GI:3861240"
/translation="MSKIQNFLLITTCISGLITLTILITPMLDLVILKHHKNNCLKEF
PWIPCTISQVMTQATITFYASSLCSIFACCTLVAVETISDEVQVWPFHSSLPKPLIKTI
SGSWSSHEGSMILMFLCLOJVSICYIECLDKMLKRSIIILISAILOFSSFIYFNS
PFHMFSEVPEKEGLINPLODIALSHIPPLLYAGVYAVETIVCAIPLAVISRDL
NMLPGYPTGKITOLSDRITSIRLKEFACGILEFTTIGISGMAVRELNGGLME
FDPENSTLPEWISGIMLHSHIITVTRKTNMOMHTIILITFLVITFSEFLRSGFI
NSIHAFPSSESRNITLYIICIGIYIIRLNNSIEKTIYNTPNMTNTHSCNMT
IKERAILIGNIFFEFLSLILICATVPTIYALNDRKPTINEEFIKNEFTIEVILIC
TIGLEPTNSSIRKHIIMLMSLITVFLINKEFSVLSITMTISFLAMNHYLLV
KTWYFRAGSKASTSMILGHPVALLAFSTYNSLSQSEMDFGVOTSKTFNAPFVT
LQNIKFRQAGKNYROIAPFMEDEHKNNITLIKPENREYIIEOKLSQSDIHSYLLVDL
YAVLSNDGKIHAKIYKPMSPFWIGVILTTAGFVIALIRNSN"
complement(7264. .12195)
/gene="Rp704"
complement(7264. .12195)
/gene="Rp704"
/translation="MAQKPNFLRKIIISAGLVATSTATYVAGSGVAMGAMOYNRTTN
AAATFEDGIEFDQACANIPVAPNSVITTAANNPITFENPGLNLSLFTDADLAVT
INDPTTIGFTINIAOAKPFNFVPAKGLINITGOGITVQEAENTIAOALTKHVGG
AALINANDSLGISTITFAAPASVLEPNINPTQEAETIGANSKTYNGKNTING
FIOVSDNTPAGKITINIDCGILMFNSTPDAANTLNQVGSNTINRGIDGTRKLVLY
SKNGATEFNVSTTLGNGLKGIILNAAVAGLISQGVAAVAVICTDNGSAGAAPIE
VSDNGNAATISQVYAKMNVIOSANAGVTEHEIYDVGAGTNEKADSVKIIIE
NSNFGSTINFGNDTOIIVPDKILKFGIEDVNNNGNTAGVIFENAGLVASSTDPN
IAVTNINAIAPACGAVVLESGIHAEIRLNGNSIRKLADVITNGVNONALMNNA
LAAGSIOLGSAIITIGDNGVNAALOHTILANDSKIALALDANGCALANGCAIHE
LANGCTIKLTNTQNNIYVNDLITDTRKVDVDAVSSITNNQITTINGISCTYVANKT
LAOLNIGSSKTIINAGDVATNELVTEKNSVQVNHNTYLTIKTINANGQIIIVADP
LNTNTTLADGTNLGSAENPLSTHFAKTAAGSLILNVGVMVANNITTTNDANGS
LHFRSGSTISVSGTGVGOQGHKLNMLINDGTAVKFLGDTFNGGTRIEGKSIIQISN
NYTTHVESADNTGLEFVNTDPTITVPLKOGAFVGLKOVIIISGPNVFNIGVAG
IVHGIAANSISFENASIGTSLFPLSGPTPLVDLITRKVNGVITNDNAPVYAGSIDS
MINNGOIIQDKKNTIALISGSDNSTYVANTLTLSGRTTRKNQGTITLSGAPNNNGT
ITGLVLENGSPKLVQYFTTIDVNNLSIILANNVTINDVYTLTGGLAGDFDKITLIG
SYGNAVNFVDSFTSDPRSMIVATQANKCTVYLDNALVSNISLDTPIVASVRFGN
DSGAGLOGNITVSONIDFGTYNLTILMSNVILGGTAINGEIDLTNNLITFANGISTW
GDNTSISTTLNVSNGNIOVYVLAEDOVANATGGTTIKIDOMANNFEGTOAYTLIO
GGRFNGTIGAPNAYTGSNIPIKVELIRPSNDQVYTPYGAITATDPSAAVTVNINDPCKT
APVSONISRCSLESTNTAAVNNMLAKDPSDVAETPYGATATDPSAAVTVNINDPCKT
QDLISNLGTLRLTLSNAETSDVAGSATGAVSSDEAEVSTGWAKFPYINAEODRKGG
IAGYKAKTGVVGLDTPLASDNLMIGAAGITTYTDIRHODYKGDFTDINGLSFSLYG
SQOLVNFEPAGNAIFTLNKKVSKRSQRAEPESGKKSKQIAAGVNDNMTFGSLISYG
DYNAMPNVLTVPMAGLSYLKSSENNYKCTGYTANKRINSKFEEDRDLIVGAIVAST
VNTIDVIVPEIHSFVHVHKVKNKLSNSQSLDQGTAPFISQDPRTAKTSYINGLSANI
KSDAKMEYIGCYDPSASKYTAHQGLKRVN"
complement(12953. .13216)
/gene="Rp705"
complement(12953. .13216)
/gene="Rp705"
/translation="MPTLIIVIIIMISLSALFAETATATASSPKIHKLITAGNKR
AKTVLEVLKKKEVIGILLIGNSLINTCTTATLTLSLGNNGTIVAGVAFIT
VRAEYVKATAIVAKPEDLTKMSSTIVFLKLRPIINADYITKJECFIRINLNO
ISGTEEVGVIEYHOGGVYKSERNNVGLIIRNMTVSEIMTHRSNIIALNIDLPY
ELIKITLISGAHRIPLMDNRNDIIGILNKLKALYENNDKTKVDINTLTPEPM
FTIDNALVVDQLHAERFRRNNHAGVDEYGLTGLITLEDVIEIYVAGTIDDEHRLNN
EIIQESMTERTIKTITRIDINRELDMNLSDEPANTAGLIHKIARIIPNOGEVITKE
NFKIILKRIANKIDSVKITVLPMTERTISSE"
complement(13563. .14507)
/gene="Rp706"
complement(13563. .14507)
/gene="Rp706"
/codon_start=1
/transl_table=1
/transl_table=1
/product="unknown"
/protein_id="CAAI5141.1"
/db_xref="GI:3861242"
/translation="MKIRQTAQPVITIGISGPELTDARELFEENHNPDIILFRNRIRK
NEKGEQDKAALIKLVIDIKVGLDNTIINADGGVYKRLIAPITDADAPQOTFLEIA
TCXYNSYTIKLELREVGINDFAPVADLIHDGCAIISDRSKSFKEPEIVVPLFSLA
```


Db	53911	GGCACTAAATATTTCTATATGGGACTACTCAAGAGAATGTTATATATATACACCTTGATGAT	53970
QY	1338	GAGCAAAAAGTCGAAGGTCGCTATTGATGTAGTTCGCCGCCGCTTCGGAATAATCA	1397
Db	53971	GAACCTTAAACAAAGCAAGCAGCTGATGTATAGTCAAAAGTTGGAAATCC	54030
QY	1398	GGAGGAGCTTAAATCCAA	1415
Db	54031	TCCTAGTGGCTTTGACCA	54048
RESULT	38		
LOCUS	AE008663	12723 bp	DNA linear BCT 14-SEP-2001
DEFINITION	Rickettsia conorii Malish 7, section 95 of 114 of the complete genome.		
ACCESSION	AE008663	AE006914	
VERSION	AE008663.1	GI:15620258	
KEYWORDS			
SOURCE	Rickettsia conorii.		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.		
REFERENCE	1 (sites)		
AUTHORS	Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P. E., Raoult, D. and Claverie, J. M.		
TITLE	Selfish DNA in protein-coding genes of Rickettsia		
JOURNAL	Science 290 (5490), 347-350 (2000)		
MEDLINE	20485642		
PUBMED	11030655		
REFERENCE	2 (sites)		
AUTHORS	Ogata, H., Audic, S. and Claverie, J. M.		
TITLE	Selfish DNA and the origin of genes		
JOURNAL	Science 291 (5502), 252-253 (2001)		
MEDLINE	20485642		
PUBMED	11030655		
REFERENCE	4 (bases 1 to 12723)		
AUTHORS	Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and Raoult, D.		
TITLE	Mechanisms of evolution in Rickettsia conorii and R. prowazekii		
JOURNAL	Science 293 (5537), 2093-2098 (2001)		
MEDLINE	21442074		
PUBMED	11557893		
REFERENCE	4 (bases 1 to 12723)		
AUTHORS	Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and Raoult, D.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CRS UMR 6020, Universite de la Mediterranee, 27 Boulevard Jean Moulin, Marseille Cedex 05 13385, France		
COMMENT	A public version of R. conorii genome database is accessible at http://igs-server.crs-mrs.fr/ . The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.		
FEATURES	location/Qualifiers		
source	1. .12723		
	/organism="Rickettsia conorii"		
	/strain="Malish 7"		
	/db_xref="taxon:781"		
	214. .489		
gene	/gene="RC1133"		
CDS	214. .489		
	/gene="RC1133"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AA03671.1"		
	/db_xref="GI:15620259"		
	/translation="MTRNGIKAIKILPIPIPEYLLRYMELAKEPICLKRLKIFHGF		
	DDENYIPNALASHKLIINFSLLKHFILKPLIDISTDPNPIYPMWS		
	complement(780. .1544)		
	/gene="RC1134"		
gene			
Db	53911	GGCACTAAATATTTCTATATGGGACTACTCAAGAGAATGTTATATATATACACCTTGATGAT	53970
QY	1338	GAGCAAAAAGTCGAAGGTCGCTATTGATGTAGTTCGCCGCCGCTTCGGAATAATCA	1397
Db	53971	GAACCTTAAACAAAGCAAGCAGCTGATGTATAGTCAAAAGTTGGAAATCC	54030
QY	1398	GGAGGAGCTTAAATCCAA	1415
Db	54031	TCCTAGTGGCTTTGACCA	54048
RESULT	38		
LOCUS	AE008663	12723 bp	DNA linear BCT 14-SEP-2001
DEFINITION	Rickettsia conorii Malish 7, section 95 of 114 of the complete genome.		
ACCESSION	AE008663	AE006914	
VERSION	AE008663.1	GI:15620258	
KEYWORDS			
SOURCE	Rickettsia conorii.		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.		
REFERENCE	1 (sites)		
AUTHORS	Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P. E., Raoult, D. and Claverie, J. M.		
TITLE	Selfish DNA in protein-coding genes of Rickettsia		
JOURNAL	Science 290 (5490), 347-350 (2000)		
MEDLINE	20485642		
PUBMED	11030655		
REFERENCE	2 (sites)		
AUTHORS	Ogata, H., Audic, S. and Claverie, J. M.		
TITLE	Selfish DNA and the origin of genes		
JOURNAL	Science 291 (5502), 252-253 (2001)		
MEDLINE	20485642		
PUBMED	11030655		
REFERENCE	4 (bases 1 to 12723)		
AUTHORS	Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and Raoult, D.		
TITLE	Mechanisms of evolution in Rickettsia conorii and R. prowazekii		
JOURNAL	Science 293 (5537), 2093-2098 (2001)		
MEDLINE	21442074		
PUBMED	11557893		
REFERENCE	4 (bases 1 to 12723)		
AUTHORS	Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and Raoult, D.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CRS UMR 6020, Universite de la Mediterranee, 27 Boulevard Jean Moulin, Marseille Cedex 05 13385, France		
COMMENT	A public version of R. conorii genome database is accessible at http://igs-server.crs-mrs.fr/ . The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.		
FEATURES	location/Qualifiers		
source	1. .12723		
	/organism="Rickettsia conorii"		
	/strain="Malish 7"		
	/db_xref="taxon:781"		
	214. .489		

gene
CDS
/gene="RC1139"
complement(5208..5582)
/gene="RC1139"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL03677.1"
/db_xref="GI:15620265"
/translation="MMHHKRLDKMLDQPGHGDGSDNLTNVAVKAEIESGNETKITN
KEIFDIDHTIERTHKEPAHYHDYRFLKTVNDNFINKNESMELKMFNSDYSKLA
TELEFSVTRMEIKEMTINHPC"
6026..6202
/gene="RC1140"
6026..6202
/gene="RC1140"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL03678.1"
/db_xref="GI:15620266"
/translation="MLEFYIAGDLKSPAFILKCGADIVATMGAMLAYEHYDANS
QESPIMIGEAQAAI"
complement(6255..7154)
/gene="t1yc"
/note="RC1141"
complement(6255..7154)
/gene="t1yc"
/codon_start=1
/transl_table=1
/product="hemolysin C"
/protein_id="AAL03679.1"
/db_xref="GI:15620267"
/translation="MLKSSKKEDSSKKNNKKLIPTVRKLESPITKNEFRKKTIDNEF
GVIRKLINSQKMTLDERNLTLANLEEDKTEDIWPERSIAIKLTNNLESESI
KLEVPHTLLYDGLDNLVGFIIKDLFALATKQNRKLKLRKHLIAAPSKLSD
LLAMRRERHIAIVDEYGGTGLVTEDEIEIYGRIDDEHQDLSDFKVIYNS
TIIANAREVELEIEIGERLHNDYDEDTIGGVLFTRVSVPAIGTRIDISENIEI
VTDAIPRLKOVKIRLKNGLNGO"
complement(7171..7674)
/gene="RC1142"
complement(7171..7674)
/gene="RC1142"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL03680.1"
/db_xref="GI:15620268"
/translation="MINVEIVRNNKREHKQINSLIKITTONILLRFDNFSKIKOP
ELSLITLAAEILTLKOFNIEKATNVLSFNSMLNODLSKLEPLGDSGYMLGD
IAPCYEVIYNESCQHKTFENHFILHLIHLILGFDHONDTEANIMENLEIELISY
FGIRPEY"
complement(7701..7904)
/gene="RC1143"
complement(7701..7904)
/gene="RC1143"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL03681.1"
/db_xref="GI:15620269"
/translation="MOVIPLTSNIDKCYPCEAAYWISGKKGKAMTDRIPTYSKLRKS
KINVISNEDIOLIKYALKLOLSL"
complement(7993..8223)
/gene="RC1144"
complement(7993..8223)
/gene="RC1144"
/codon_start=1

Query Match 11.9% Score 194.4; DB 1; Length 12723;
Best local Similarity 48.2%; Pred. No. 6,4e-36;
Matches 645; Conservative 0; Mismatches 681; Indels 12; Gaps 3;

81 AATTCGCGCTCTTCTTGCGCGATACATCAAGAGCTAAAGAAAGTCTCGCAATG 140

Db 3219 AATTCAGAGCGGCAATTTGGCGCTGCGATATACGAAGCTGGAAAGTTATTCGGATG 3278
OY 141 TTCTAATGCTTCTGTATTAACATTAACATACGAGGTGTAACGGAACAAAGACACT 200
Db 3279 ACTACCTTAATGTTGTTTATTAATTAATCAAAATGTTTACGAATCTTAAGAAATAGT 3338
OY 201 CTATTGCGGAGCTCCGTTCTGTGTCAGAGGCAATACCTTTATCAAGTTTGGCTT 260
Db 3339 ATTTTAATCT-----CCGAGATAGTCGAGAAATAGCAGGTTTGTCAAGCTTAATGCG 3392
OY 261 GTTGTCCTGCTGCTATTAATCTTATGCTTATTTGCAAGCACTAATATTTAATG 320
Db 3393 GTTACGCCCTGCCGCTGCTTATTCGTGATTTATTTGCAAAATATATATATCTTACT 3452
OY 321 AAGCAGGCGTTATTTTATGACAGTGGAAAGCCCTTTTAAATTTCTTGGCCCTGCGG 380
Db 3453 TTGAAAGAGATCTTTTATTAATTAAGTCATTTTATACAGCTTTTGTATTATTCACT 3512
OY 381 ACTGTAATTTATCCGCTACGCGATGTTTACATCCTACAGAAATTTGCTGACCGTTACAG 440
Db 3513 TTTGTAATCTATCCATAATATTCATATTTTCAATGACATCCCAATATCTAGCTACTG 3572
OY 441 GCCATCCATACCTCCAGAGATTGCTAGGACGCTTGGCATCTTAAGAACTGACATTTGCT 500
Db 3573 ATGGAACGTTATCTCTATTTTAAGTGACATCTCATTAAGTAAATTTGGGATTATATA 3632
OY 501 GCATTTTATGACTTGCGTAACATAGGGAAGCGCATGCTATCTATATGTTCTGGGGA 560
Db 3633 CATATTTATAGCTTATCCGAGCTTGGCCTAATATTTTATGATTAATTTATTTGGCAG 3692
OY 561 TTTGCTAATGAATTAACAAAATCCAGCAAGAAAGCGTTTCTAGCGTCTTTGCGTATC 620
Db 3693 TTGCGTAAGCACTACATCACTACGCAAGAAAGAAAGATTTTAATCACTCTTTTCGCTA 3752
OY 621 GGAGCTAATATTTCTTACTACTCTGCTGCTGCAATGTTGGGCTCAAGTTGAGA 680
Db 3753 TTGCGCAACTCTCTTATATATTAATGCGCTTTTAATGATCAATCATACGGAAGAT 3812
OY 681 GCTTCGCTTCTGAAGGTGTAGATCCTTGGGGAATTTCTTACGCTTTTGATGGCTATG 740
Db 3813 ACTATTAATGAAGATTTATGAGTATTTGCGATGAGTAATACATTAATGTCAGATGCC 3872
OY 741 ACTATTTGATCTGCACTTCTTATGCGCACTTACGCTGATGATTAAGAACTATTTG 800
Db 3873 ACGACAATTTGGCAATTTGCTATTAATTTGTTGTTGTTAGATTTTATGCAAG 3932
OY 801 ACCGATCCCGCTCTATATCCAGAAAGATGCAAAAGGGAAGGAAAGGCTCTAAACT 860
Db 3933 AACGTTTACTAATCCGCTATTTATGCTAAAGCAAAAGCGGTCGCTCAACTTCAGAA 3992
OY 861 AAAATGAATATGAAGATAGCTTCTCTATCTGTATAGATCTCTATATTTCTTTATATA 920
Db 3993 CSTATGCGGCTAATTAAGGCTTTAAATATATTTGGAAATCAAAATATTTATGACTACT 4052
OY 921 ACTCTCTGTTATTTGCCATGCTATTTGCACTTAATTAATGGAAGTACTGGAAGACT 980
Db 4053 TTAATTTGTTGGCAGCTTGGGATTTGCTATTAATTAAGTGAAGCCGTAAGGAAGCA 4112
OY 981 CAGCTGAACCTCAATATCTATATGATGATGATAGGATGATGGGGAAGCTTCTCC 1040
Db 4113 AAAATTAAGAAATTAATTCGCACTGTAATTAATTAATTAATTAATTAATTAATTAAT 4172
OY 1041 TTCTGAGCTGAGGTAGTATCCGTACTATATCATGATTTGTTGGTGAAGCTCATTTGCT 1100
Db 4173 CTTCGACAGCGCTGG---CAATTAATGTTATGACATCATGCAATTAATTAATTTAGCT 4229
OY 1101 AATTTGATGGTTAAGTGAAGCCCTAGACATCTCTGATGATGTTCTCTAACAGTATC 1160
Db 4230 ATGCAACAAGTGTGTTGACACGACGATTAATTTACCGGTTATATATATGATGACGCAAT 4289
OY 1161 GTTTCTTCGCTCTGTTGATCTTTAGAAACCAAGCTTCTGGGCTGCTGCTGTTGTTGGT 1220

gene 16105..17094
/gene="P516.4"
/codon_start=1
/product="hypothetical protein; 16105-17094"
/protein_id="AAG52433.1"
/db_xref="GI:1324978"
/translation="MGKVSVPRTWFLFVVDIVYCRPTLSCGDPDPCATNTNREOKLR
SQQTIVLINGYSEYRIRIPLOTTIVASYSSTSIYSLVLMGNSTPTDOLLDTLQNTLT
YSPGSAISLIQOSSSSLNARFLPRSSVDYTRAVLLCDDVEDDQSLSEFAVSKNPN
DLVQYFVNSHGFDLQGEKMYTTHPDRYSYLTNPFMMKMDIYLEYSCKGVNEEM
RNIIVHRRNCEIDILNMFVAADRLRAGPILMGVEYRMDGADNEVEYERVDGLSTR
RYEHKRRRNCRCIFREHFVYKMPKMYISGKVYNSGEGCLCKKAKLVFCORD"
19371..21989
/gene="P516.5"
/note="identical to adenine nucleotide translocase
GB:249227 [Arabidopsis thaliana] (FEBS Lett. 374 (3),
351-355 (1995))"
join(19371..20262,20466..20720,20810..20953,21035..21382
21462..21989)
/gene="P516.5"
join(19474..20262,20466..20720,20810..20953,21035..21382
21462..21989)
/gene="P516.5"
/codon_start=1
/product="adenine nucleotide translocase; 19474-21800"
/protein_id="AAG52434.1"
/db_xref="GI:1324978"
/translation="MEAVIOTRGLSLPTPKPIGRSOLPSHGKLRFLFAKPRRLHG
LSLSENGHKKQTFPEPTLHGISISHKESTFEICAEAAAGDGAVFEGESAAAYVAS
PKIPEVAVATLKIKITPLIGMFRCILPNTLITLIDPKDYLIVAKGSAARITPTKWN
LPMAGFMILYIKRLSNLSKRALFTVYPIPTITGATGFVNPISNTIHEDALADKL
LTLTLEGREGLPAILRLWSEFCLEFYMAELGWSVVAVLFGVGNITTVDEAKREPL
FTLGNALVATLESKGRTRYKFSNLKNGIQRVDMAVSLAAMSIVYGMOLACILYMW
NRYVPLPISFKGKERPKMKTMESLKFVSSPYIRDLATLVAVGISLTVLEWTMSK
LKAQSPSPREYSAPMGDESTCTGVATFTFMILSOYFKNYGVAKITPTLTYLTD
APFSLIFEGREPAPIYAKTGMTPLAAYVYGLOVIFSKAKYSYLPDCKEMAYPLID
EPTKXKRAIDVYCNPLKSGCALIQPFMILSRSLANSTPYLGMITLLVITYANLAA
NASLREQFSLSEELERKEMERASSVKIPVVSODESGNSLGESSSPSPERSAPTNL
"
gene 22471..23865
/gene="P516.6"
/note="contains Pfam profile: PF00916 sulfate transporter
family"
22471..>23865
/gene="P516.6"
22471..23865
/gene="P516.6"
/codon_start=1
/product="putative sulfate transporter; 22471-23865"
/protein_id="AAG52436.1"
/db_xref="GI:1324980"
/translation="METTITFPLPDGRSCGWLRRRLRLKPNLSELSGAAGDGTFI
PIVLTVALISNDLSTTLITTFGEVNIATGLIPDIPMPQPKMSIAVAVSSPHRTPS
QIAAAGASTATLILLIGATGANSFLYNIITPLEYVNGVLSQGLQFAITAKIVRNTD
TATRLPSSPRLWIDGLILALALFLITSTSGGNREARDGGLATSSNESQSLRA
RATPLRLSPSLAIVFALGLVLCFTRDPSIFDKLFGSKFHLIRLISDMDKIGSRA
AIQPIQLSVLNSVIAVCKLSNDLPKELASITSGVMNLIIGCFAMPGCHAGG
LAGQVREGKRGDSVIFLIGIKLIVLGVNFSVFLISQFPIGLIGVLLFAGIEILAM
ASQDNKRSKRDYILVCAAVSMTGSSALGFCGVVILLLKRLRLDSSSVTLPERSS
DSSQVDSKAPRDV"
24302..25416
/gene="P516.7"
/note="contains Pfam profile: PF00671 Iron/Ascorbate
oxidoreductase family"
join(<24302..24678,24752..25079,25159..>25416)
/gene="P516.7"
join(24302..24678,24752..25079,25159..25416)
/gene="P516.7"
/codon_start=1
/product="putative oxidoreductase; 24302-25416"
/protein_id="AAG52438.1"
/db_xref="GI:1324982"
/translation="MDSHEEQCLPIIDFSSDKLVYRGTSMTITTRDQVRRAMEGQGFV

Query Match	Best Local Similarity	8.5%; Score 139.8; DB 8; Length 132990;
Matches	261; Conservative	0; Mismatches 202; Indels 0; Gaps 0;
QY 121	TAAGAAAGTTCTGCCAATGTTCTCTAAATGTTCTCTGTAATACATTTTAACATACGCTGT	180
Db 19811	TGAAAAGATTTATTCCTTTAGGATTTGATGTTCTTTGTAATCTTTCAATACCAATTC	19870
QY 181	TACGGATACAAACACACTCTATTGGTGGAGCTCCGTTCTCGTGCAGACGAATAC	240
Db 19871	TGAGGAGTACAAAGATGCTTTGGTGGAGCGCGCAAGAGATTCCTGATATTAAC	19930
QY 241	CTTTCATCAAGTTTGGCTGTTGTGCCCTGCGTATTTATCTTTATGTTATTAATGCA	300
Db 19931	CTTTCTTGAAGACTTGGGTGGAATCTTCTCATGSCATTGGGTTTATGCTCTCTACACTA	19990
QY 301	AGCTAAGTAATATTTTAAGTAAGACAGGCGCTTATTTTATGCACTGGGAACGCCCTTTTAA	360
Db 19991	AACCTCCCAATGTTCTCTCCACAGAGGCGCTGTTTATACATGTTATGTCCCTTCATCA	20050
QY 361	TTTTCTTTGGCTGTTCCCGACTGAATTTATCCGCTACGGATTTTACATCTACAG	420
Db 20051	TCTACTTTGGGGCCCTTGGTTGCGCATGACCCCTCAGAACATATATACCCGGAG	20110
QY 421	AATTTCTGACCGTTTACAGGCACTCAACCCAGATGGAATGGCTAGACATCGTTGCACT	480
Db 20111	CTCTGCGCATATAGCTCTCTTACACCCCTCGGCCCAAGATTCATGSGTCTCTATGCAATAT	20170
QY 481	TAAGAACCTGCACATTTGCTGCATTTATATGACTTGTGAACATATAGGGAACCGTCATGC	540
Db 20171	TGCGGATTTGGAGTTTCTGTTGTTTATGTTATGTTATGCTAGACCTTGGGGTATGCTGTG	20230
QY 541	TATCTCTAATGTTCTGGGGATTTGCTAATGAAATTACAAAAT	583
Db 20231	TCTCAGTTCTCTCTGGGCGTTTGCTAATACAGTGACTAAACT	20273
RESULT 40		
AC126791		
LOCUS	AC126791	
DEFINITION	Medicago truncatula clone mtl1-7k2, WORKING DRAFT SEQUENCE, 28	
ACCESSION	AC126791	
VERSION	AC126791.11	GI:22450616
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	barrel medic.	
ORGANISM	Medicago truncatula	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.	
REFERENCE	1 (bases 1 to 21975)	
AUTHORS	Shuall,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.	
TITLE	Medicago truncatula BAC Clone mtl1-7k2	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 21975)	
AUTHORS	Shuall,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (09-JUN-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	
REFERENCE	3 (bases 1 to 21975)	
AUTHORS	Shuall,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-AUG-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	

COMMENT

On Aug 23, 2002 this sequence version replaced g1:22417350.
 ----- Genome Center
 Center: Department of Chemistry And Biochemistry
 The University of Oklahoma
 Center code:UOKNOR

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 2033: contig of 2033 bp in length
* 2034 2133: gap of unknown length
* 2134 4316: contig of 2183 bp in length
* 4317 4416: gap of unknown length
* 4417 7180: contig of 2764 bp in length
* 7181 7280: gap of unknown length
* 7281 9824: contig of 2544 bp in length
* 9825 9924: gap of unknown length
* 12804 12903: contig of 2879 bp in length
* 12904 15185: contig of 2282 bp in length
* 15186 15285: gap of unknown length
* 15286 17545: contig of 2260 bp in length
* 17546 20257: gap of unknown length
* 20258 20357: gap of unknown length
* 20358 23799: contig of 3342 bp in length
* 23700 26276: contig of 2477 bp in length
* 23800 26376: gap of unknown length
* 26277 29259: contig of 2883 bp in length
* 29260 29359: gap of unknown length
* 29360 32359: contig of 3000 bp in length
* 32360 32459: gap of unknown length
* 32460 36613: contig of 4154 bp in length
* 36614 39318: gap of unknown length
* 39319 39418: gap of unknown length
* 39419 41836: contig of 2418 bp in length
* 41837 41936: gap of unknown length
* 41937 45761: contig of 3825 bp in length
* 45762 45861: gap of unknown length
* 45862 51599: contig of 5738 bp in length
* 51600 51699: gap of unknown length
* 51700 58201: contig of 6502 bp in length
* 58202 58301: gap of unknown length
* 58302 64490: contig of 6189 bp in length
* 64491 64590: gap of unknown length
* 64591 70460: contig of 5870 bp in length
* 70461 70560: gap of unknown length
* 70561 75838: contig of 5278 bp in length
* 75839 75938: gap of unknown length
* 75939 82428: contig of 6490 bp in length
* 82429 82528: gap of unknown length
* 82529 87785: contig of 5257 bp in length
* 87786 87885: gap of unknown length
* 87886 95365: contig of 7480 bp in length
* 95366 95465: gap of unknown length
* 95466 102528: contig of 7063 bp in length
* 102529 124627: gap of unknown length
* 124628 124527: contig of 21799 bp in length
* 124528 169060: contig of 44533 bp in length
* 169061 217975: gap of unknown length
* 169161 217975: contig of 48815 bp in length.
  
```

FEATURES

location/Qualifiers
 1..217975
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"

/clone="mchl-7k2"
 /clone_lib="Medicago truncatula BAC library H1"
 BASE COUNT 71744 a 35163 c 35694 g 72650 t 2724 others
 ORIGIN

Query Match 8.4% Score 138; DB 2; Length 217975;
 Best Local Similarity 59.2%; Pred. No. 1.5e-22;
 Matches 271; Conservative 0; Mismatches 185; Indels 2; Gaps 2;

```

QY 120 CTAAGAAGATGTCGCCAATGTCCTATGTCCTGCTATACATTAACATATACGCGTG 179
Db 50970 CTCAGAAGATTAACACCTGAGATGATGTTTGTATGTTCAATTAACAAATT 51029
QY 180 TTACGCGATACAAAAGACACTATTGTCGAGCTCCTCGTTCGTGAGAGCAATA 239
Db 51030 CTAAGAGATACAAAAGATGTTCTGTTGTGACGTGACGAGAGATGAGCATATA 51089
QY 240 CCTTTCATCAAGTTTGGCTGTTGTCCTCCCTGCTGCTATTAATCTTAATGCA 299
Db 51090 CCTTTTGAAGAACATGGGTGAATCTTCATGCTATGCTTATGTTGTGTACACA 51149
QY 300 AAGCTA-AGTAATATTAATTAAGTACAGGCTTATTTATGCACTGGAGCCCTTTT 358
Db 51150 AAGTATTCATATGTTGTCACAAACAGGCTTTTATCTGTATGTCCTTTAT 51209
QY 359 AATTTCTTCCCTGTCCTCCGACGTAAATTAATCCGCTAGCATGTTTACATCCAT 418
Db 51210 TGCCTTCTTGGGCTTTGGGCTTTGTTGTATCTTATGCAACTATATACCTCA 51269
QY 419 AGAATTCGTGACCGTTTACAGGCCATCTACCTCAGAGATGCTAGGACTGTGCAT 478
Db 51270 GGCATTTTG-TGATTAACCTTCTTAATATCTTGACCTCGGTTCTGTGCGCTGCTAT 51328
QY 479 CTTAAGAACTGACATTTGCTGATTTATGATTAATGCTGCTAATGAGGAGCGCAT 538
Db 51329 TATGAGGATTTGAGATTTGCTGTTCTATGTCATGAGCGAATGAGGAGGTGTG 51388
QY 539 GCTATCTCTAATGTTCTGCGGATTTGCTAATGAATTA 576
Db 51389 GATTCGTGCTGTTTGGGGGCTGCGCAATCAGGTAA 51426
  
```

Search completed: January 21, 2003, 19:36:30
 Job time : 5227 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 18:03:28 : Search time 59 Seconds
(without alignments)
8508.981 Million cell updates/sec

Title: US-09-869-433-1

Perfect score: 1637
Sequence: 1 gaataaaacatactacgaag.....gttcacccactcgtcttt 1637

Scoring table: IDENTITY_NUC
Gap 10.0 , Gape 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCrus.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.4	2.9	7218	1 US-08-232-463-14	Sequence 14, Appl
2	38.8	2.4	2718	4 US-09-074-658-69	Sequence 69, Appl
3	38.4	2.3	1269	4 US-09-134-001C-1211	Sequence 1211, Ap
4	38	2.3	1017	4 US-09-134-001C-2318	Sequence 2318, Ap
5	35.8	2.2	17949	4 US-09-087-465-3	Sequence 3, Appl
6	35.4	2.2	1690	1 US-08-276-452A-24	Sequence 24, Appl
7	35.4	2.1	1690	2 US-08-798-744-24	Sequence 24, Appl
8	34.6	2.1	1647	1 US-08-198-446B-9	Sequence 9, Appl
9	34.6	2.1	1647	2 US-08-870-693-9	Sequence 9, Appl
10	34.4	2.1	729	4 US-09-134-001C-1161	Sequence 1161, Ap
11	33.8	2.1	658	4 US-08-998-416-595	Sequence 595, App
12	33.6	2.1	3946	4 US-09-453-702B-103	Sequence 103, App
13	33.4	2.0	405	4 US-09-134-001C-980	Sequence 980, App
14	33.4	2.0	585	4 US-09-222-939-30	Sequence 30, Appl
15	33.4	2.0	2590	3 US-08-714-918-78	Sequence 78, Appl
16	33.4	2.0	2590	4 US-09-265-315-78	Sequence 78, Appl
17	33.4	2.0	2590	4 US-09-265-315-78	Sequence 78, Appl
18	33.4	2.0	2590	4 US-09-265-315-78	Sequence 78, Appl
19	33.4	2.0	2706	4 US-09-463-238-2	Sequence 2, Appl
20	33.4	2.0	2806	4 US-09-463-238-11	Sequence 11, Appl
21	33.4	2.0	3483	4 US-09-130-491-3	Sequence 3, Appl
22	33.4	2.0	17410	1 US-07-841-646-3	Sequence 3, Appl
23	33.4	2.0	17410	1 US-08-147-023-3	Sequence 3, Appl
24	33.4	2.0	17410	1 US-08-447-570-3	Sequence 3, Appl
25	33.4	2.0	17410	2 US-08-449-700-3	Sequence 3, Appl
26	33.4	2.0	17410	2 US-08-449-699A-3	Sequence 3, Appl
27	33.4	2.0	17415	3 US-08-486-343A-1	Sequence 1, Appl

c	28	33.4	2.0	17415	5 PCT-US95-07349-1	Sequence 1, Appl
	29	33.2	2.0	1691	3 US-09-013-981-12	Sequence 12, Appl
	30	33.2	2.0	3989	1 US-08-327-494A-1	Sequence 1, Appl
	31	33.2	2.0	3989	1 US-08-327-494A-3	Sequence 3, Appl
	32	33.2	2.0	3989	5 PCT-US95-13659-1	Sequence 1, Appl
	33	33.2	2.0	3989	5 PCT-US95-13659-3	Sequence 3, Appl
	34	33.2	2.0	5163	3 US-08-700-651-1	Sequence 1, Appl
	35	33.2	2.0	5163	3 US-08-928-361B-4	Sequence 4, Appl
	36	33.2	2.0	5318	3 US-08-700-651-2	Sequence 2, Appl
	37	33.2	2.0	5318	3 US-08-928-361B-3	Sequence 3, Appl
	38	33	2.0	200	4 US-09-020-956-140	Sequence 140, App
	39	33	2.0	200	4 US-09-030-607-140	Sequence 140, App
	40	33	2.0	200	4 US-09-605-785-140	Sequence 140, App
	41	33	2.0	200	4 US-09-429-313-140	Sequence 140, App
	42	33	2.0	200	4 US-09-352-616A-140	Sequence 140, App
	43	33	2.0	200	4 US-09-232-149A-140	Sequence 140, App
	44	33	2.0	2220	4 US-08-934-386-2	Sequence 2, Appl
	45	33	2.0	4619	2 US-08-874-186-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHREIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOULPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 2.9%, Score 47.4, DB 1, Length 7218;

Best Local Similarity 5.3%; Pred. No. 0.0028;
Matches 21; Conservative 210; Mismatches 166; Indels 0; Gaps 0;

```
OY 130 TTTCGCCAATGTTCCCTAATGTTCTTCGTATACATTTAACAATAGCGGTAGCGGATA 189
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1080 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1139
OY 130 CAAAGACACCTTATTTGGGAGCTCCTGGTTGCTGACAGGCAATACCTTTCATCA 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1140 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1199
OY 250 AGTTTGCTGTGTGCCCTGTGCTATTTATGCTATTTATGCAAGCAATCA 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1200 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1259
OY 310 ATATTTAAGTAGAGGAGCTTATTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1260 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1319
OY 370 CCCGTGTCGACGTAAATTTATCCGCTACGCGATGTTTACATCCATACAGATTTGCTG 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1320 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1379
OY 430 ACCGTTACAGGCAATCCACCTCCAGATGCTAGAGCTGCTGCTGCTGCTGCTGCTGCTG 489
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1380 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1439
OY 490 GGACATTTGCTGCTGCTTATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1440 AATCTCTCTATCTCTTAACTACTGATGATG 1476
```

RESULT 2

US-09-074-658-69/c
; Sequence 69, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2718 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: Linear

US-09-074-658-69

Query Match 2.4%; Score 38.8; DB 4; Length 2718;
Best Local Similarity 57.4%; Pred. No. 0.44;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
OY 1079 TGTGTGCTGACGTCAATGCTAAATTTGGATGCTTAAGTGGAGCCCTAGTCACTCTGT 1138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1231 TATGTGCTGATCATCAATCAAGATTTGGCTGTGCCAGTAAGCGGTTGCCATTAATCT 1172
OY 1139 CATGCTTCTCTACAGGATGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1171 TGGGCTCATCTCATGACATGCGCTTCATGACCTTTTGTGCTATCAACATGACTTC 1112
OY 1199 TG 1200
      : :
Db 1111 TG 1110
```

RESULT 3

US-09-134-001C-1211
; Sequence 1211, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1211
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1211

Query Match 2.3%; Score 38.4; DB 4; Length 1269;
Best Local Similarity 47.5%; Pred. No. 0.41;
Matches 114; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
OY 1019 TGAATTCATGGGAGCACTTCCTTCGACGCGGCTAGTATCCGTACTTATCATGATTT 1078
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 885 TAAATTAACAAAATTTATACCTACATTTCTTTACTATTGTACAAATCACTACTATT 944
OY 1079 TGTGTGCTGACGTCAATGCTAAATTTGGATGCTTAAGTGGAGCCCTAGTCACTCTGT 1138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 945 TGTACTGTTATTCAAAATTCATATAGCATATATCATTTGCTTTATTCGGTTTGG 1004
OY 1139 CATGCTTCTCTACAGGATGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1005 TACATTTGGTACACACCGCTTACTTAATAGCAAAATATCTTAAGTCAAAAAGAGACC 1064
OY 1199 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1258
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1065 ACTTCTTGCAAGTACGTTAGCAGCTTCTATTTCATATGTTGCTAATTTCTTGGTGAAT 1124
```

RESULT 4

US-09-134-001C-2318
; Sequence 2318, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13

RESULT 9
 US-08-870-693-9/c
 Sequence 9, Application US/08870693
 Patent No. 5866338
 GENERAL INFORMATION:
 APPLICANT: Hartwell, Leland H.
 APPLICANT: Weinert, Ted A.
 APPLICANT: Pilon, Sharon E.
 APPLICANT: Groudine, Mark T.
 TITLE OF INVENTION: Cell Cycle Checkpoint Genes
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
 STREET: 1420 Fifth Ave., Suite 2800
 CITY: Seattle
 STATE: WA
 COUNTRY: USA
 ZIP: 98101-2347
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

```

RESULT 10
US-09-134-001C-1161/c
: Sequence 1161, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1161
: LENGTH: 729
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1161

```

Query Match	2.1%	Score	34.4	DB	4	Length	729
Best Local Similarity	47.3%	Pred.	No. 4.1				
Matches 104; Conservative	0	Mismatches	116	Indels	0	Gaps	0

QY	744	ATTGATATGGACATTGGTTCTTAATGACCGAGTACGAGGATCATTAAGAAGCTTTGACC	803
QY	603	ACTATACCTATTCGTTAATCATCTCTTTGGAACTGGAATCATCTCAAAAATTGAGTATATA	5444
QY	804	GATCCCTCGCTCTATTAATCCAGAAANAATGCAAAAGGGGAAAAGCTGCTAAACCTAAA	8635
QY	543	ATATTTATTTATTCATCATATATATATTAATATCTCAAAATTCATCAAGTGTGGATTATATAG	4848
QY	864	ATGAATATGAAAGATAGCTTCCTCTATCTTGATGATGATCTCCTTATATTTCTTTATTAAT	9232
QY	483	ATAAGCATAGTCGATTAACCTTCCTCTAAATCCTCAAACTCTACATATATATATTTCTCA	4242
QY	924	CTCTTGCTTATTTGCCATAGATTTGCAATTAACCTAAATG	963
QY	423	TGATGCGCTTATGCGCAGCTCTTTGTGTTATTCCTTGATG	384

```

US-08-998-416-595/C
: Sequence 595, Application US/08998416
: Patent No. 6239264
:
: GENERAL INFORMATION:
: APPLICANT: Philippsen, Peter
: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jurgen
: APPLICANT: Knechtle, Philipp
: APPLICANT: Robischung, Corinne
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
: TITLE OF INVENTION: AND USUS THEREOF
: NUMBER OF SEQUENCES: 1152
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6239264artis Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: No. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 0016/97
: FILING DATE: 31-DEC-1996
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8567
: TELEFAX: 919-541-8689
:
: INFORMATION FOR SEQ ID NO: 595:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 658 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: PAG1408RP
:
: US-08-998-416-595

```

Query Match	2.1%	Score 33.8;	DB 4;	Length 658;
Best Local Similarity	49.7%;	Pred. No. 5.8;		
Matches 86;	Conservative 0;	Mismatches 87;	Indels 0;	Gaps 0;

Qy	2	AAATAAAAAACTATCAGAAATGAAAATTTTCACAGGGTAAATATGACAAAAAC	61
Db	611	ATAAAAATAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA	552
Qy	62	CGAGAAAAACCTTTGGAAATTCGCGCTCTTCTGTGGCCGATACATCAACGAGCT	121
Db	551	ATTAATATTAATATAGAAATAGAAATAGCATATATATATATATGTTCCAAATATCAATCTATAT	492
Qy	122	AAAGAAAGTCTTGCCAAATGTTCCATATGTTCTTCTGTATTACATTTACATATA	174
Db	491	AATGATATAAGTCACCATTAATATATACATATACATATGATATATATATATATATATA	439

```

12 RESULT 12
13 US-09-453-702B-103/C
14 : Sequence 103: Application US/09453702B
15 : Patent No. 6365723
16 :
17 : GENERAL INFORMATION:
18 :
19 : APPLICANT: Blattner, Frederick R.
20 : Burtland, Valerie
21 : Perna, Nicole T.
22 : Plunkett, Guy
23 : Welch, Rod
24 :
25 : TITLE OF INVENTION: NO. 6365723el Sequences of E. coli 0157
26 :
27 : NUMBER OF SEQUENCES: 265
28 :
29 : CORRESPONDENCE ADDRESS:
30 : ADDRESSEE: Quarles & Brady
31 : STREET: 1 South Pinckney Street
32 : CITY: Madison
33 : STATE: WI
34 : COUNTRY: US
35 : ZIP: 53701-2113
36 :
37 : COMPUTER READABLE FORM:
38 : MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
39 :
40 : COMPUTER: IBM PC compatible
41 : OPERATING SYSTEM: PC-DOS/MS-DOS
42 : SOFTWARE: Word Perfect 8.0
43 :
44 : CURRENT APPLICATION DATA:
45 : APPLICATION NUMBER: US/09/453,702B
46 : FILING DATE: 03-Dec-1999
47 : CLASSIFICATION: <Unknown>
48 :
49 : PRIOR APPLICATION DATA:
50 : APPLICATION NUMBER: 60/110,955
51 : FILING DATE: 04-DEC-1998
52 :
53 : ATTORNEY/AGENT INFORMATION:
54 : NAME: Seay, Nicholas J.
55 : REGISTRATION NUMBER: 27386
56 : REFERENCE/DOCKET NUMBER: 960296.95017
57 :
58 : TELECOMMUNICATION INFORMATION:
59 : TELEPHONE: (608) 251-5000
60 : TELEFAX: (608) 251-9166
61 :
62 : INFORMATION FOR SEQ ID NO: 103:
63 :
64 : SEQUENCE CHARACTERISTICS:
65 :
66 : LENGTH: 3946
67 : TYPE: nucleic acid
68 : STRANDEDNESS: double
69 : TOPOLOGY: linear
70 :
71 : MOLECULE TYPE: DNA (genomic)
72 :
73 : SEQUENCE DESCRIPTION: SEQ ID NO: 103:
74 :
75 : US-09-453-702B-103

```

[illegible]

OY 291 ATTATGCAAGCTAAGTAATATT 314
|||||
Db 1484 ATTCTGAAATTTATGCAATGTAAT 1461

RESULT 13
US-09-134-001C-980
; Sequence 980, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 980
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-980

Query Match 2.0%; Score 33.4; DB 4; Length 405;
Best Local Similarity 55.7%; Pred. No. 6.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 1415 ACAAGGTTGCTGCTATCTGTGAGATGAGACTATGACCCCTTCTTGCGATGAT 1474
|||||
Db 276 ACCAGGATTTTATAGCTGTTAGTAGAGTAAGTCTCTCTACCTATAGGCGTAT 335
OY 1475 TCTCTTTTCATCATGCTATTGTTGTTCTGCACTAGTTAAACAACTA 1529
|||||
Db 336 ACTATTTTCTTTAACTGCAATTCATGTCTCAGGCTACTAAGCAAAATATCA 350

RESULT 14
US-09-222-939-30
; Sequence 30, Application US/09222939
; Patent No. 6372448
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; APPLICANT: Guzman, Luz-Marie
; TITLE OF INVENTION: USE OF YLOF, YOEQ, YYBO, YERL, AND YSXC, ESSENTIAL BACTERIAL
; TITLE OF INVENTION: GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/111001
; CURRENT APPLICATION NUMBER: US/09/222,939
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-222-939-30

Query Match 2.0%; Score 33.4; DB 4; Length 585;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 680 AGCTTCCCTTCTGAGAGTGTAGATCCTTGGGAAATTTCTTACGCTTTGATGGCTAT 739
|||||
Db 94 AGCTTGCATCTTTAGACAGAAATAGTTAGTATATAGTAGCCTTTTGAAGCCTCT 153
OY 740 GACTATTTGATCTGAGCTTGTCTTATGCGCAGTTACTGCGATCAATTAAGAAC 794
|||||
Db 154 TGCAGTAGTTCCGCCCTTTTGTGTTGCGAATTAAGATGTAATAGTACTACTAC 208

RESULT 15
US-08-714-918-78
; Sequence 78, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,918
FILING DATE: September 13, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Walburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 2590 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-714-918-78

Query Match 2.0%; Score 33.4; DB 3; Length 2590;
Best Local Similarity 52.6%; Pred. No. 13;
Matches 70; Conservative 1; Mismatches 62; Indels 0; Gaps 0;

OY 835 AAAAGGGGAAAAAGTCTAAACCTAAATGAATGAAGATAGCTTCTATCTTGG 894
|||||
Db 1731 AAAATTGAAGAGTAGTATAGTAATATATATAAATAGCGCATATTGTCGTTA 1790
OY 895 ATAGATTCCTTATATTTCTTTTATTAATCTCTCTGTTATGCGATATTTGCATTA 954
|||||
Db 1791 TAAGATTTTTCATATGATTTTTCAAAGTATGTTGCTTTCATCTCATATTTGCTAAT 1850
OY 955 ACTAATGGAAGT 967
|||||
Db 1851 TGTAAAGCTATGT 1863

RESULT 16

```

US-09-265-315-78
: Sequence 78, Application US/09265315
: Patent No. 6187541
:
: GENERAL INFORMATION:
: APPLICANT: Benton, Bret
: APPLICANT: Lee, Vling J.
: APPLICANT: Malouin, Francois
: APPLICANT: Martin, Patrick K.
: APPLICANT: Schmid, Molly B.
: APPLICANT: Sun, Dongxu
:
: TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
: TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
: NUMBER OF SEQUENCES: 111
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: Storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Word Perfect 5.1
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/265,315
: FILING DATE: March 9, 1999
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/714,918
: FILING DATE: September 13, 1996
: APPLICATION NUMBER: 60/009,102
: FILING DATE: December 22, 1995
: APPLICATION NUMBER: 60/003,798
: FILING DATE: September 15, 1995
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 240/247
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
:
: TELEX: 67-3510
:
: INFORMATION FOR SEQ ID NO: 78:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2590 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-09-265-315-78
:
: Query Match 2.0%; Score 33.4; DB 4; Length 2590;
: Best Local Similarity 52.6%; Pred. No. 13;
: Matches 70; Conservative 1; Mismatches 62; Indels 0; Gaps 0;
:
: QY 835 AAAGGGGAAAAAAGCGTCTAAACCTAAATGAATATGAAAGATAGCTTCCTCTATCTTG 894
:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 1731 AAAATTTGAAAGGAAAGTATTGAAATTAAGTATTAAATTAATAAATGCGCATTTATTGCGTTA 1790
:
: QY 895 ATAGATTCCTTATTTCTTTTATTAATCTCTTGSTATTGCCATGATTTGCATTA 954
:      || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 1791 TAAAGTATTTTACCTATTTTCAAGTATTGTTGCTTTGCAATCATATATTTGCTAAT 1850
:
: QY 955 ACTTAATGCAAGT 967
:      ||| ||| |||
: Db 1851 TGTTAAGCTATGT 1863
:
: RESULT 17

```

```

US-09-265-78      US-09-265-78
: Sequence 78, Application US/09265315
: Patent No. 6187541
: GENERAL INFORMATION:
:   APPLICANT: Benton, Bret
:   APPLICANT: Lee, Vling J.
:   APPLICANT: Malouin, Francois
:   APPLICANT: Martin, Patrick K.
:   APPLICANT: Schmid, Molly B.
:   APPLICANT: Sun, Dongxu
:   TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
:   TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
:   NUMBER OF SEQUENCES: 111
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Lyon & Lyon
:     STREET: 633 West Fifth Street
:     STREET: Suite 4700
:     CITY: Los Angeles
:     STATE: California
:     COUNTRY: U.S.A.
:     ZIP: 90071-2066
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
:   MEDIUM TYPE: Storage
:   COMPUTER: IBM Compatible
:   OPERATING SYSTEM: IBM P.C. DOS 5.0
:   SOFTWARE: Word Perfect 5.1
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/09/265,315
:     FILING DATE: March 9, 1999
:     CLASSIFICATION: 435
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: 08/714,918
:     FILING DATE: September 13, 1996
:     APPLICATION NUMBER: 60/009,102
:     FILING DATE: December 22, 1995
:     APPLICATION NUMBER: 60/003,798
:     FILING DATE: September 15, 1995
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Warburg, Richard J.
:     REGISTRATION NUMBER: 32,327
:     REFERENCE/DOCKET NUMBER: 240/247
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE: (213) 489-1600
:     TELEFAX: (213) 955-0440
:     TELEX: 67-3510
:   INFORMATION FOR SEQ ID NO: 78:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 2590 base pairs
:       TYPE: nucleic acid
:       STRANDEDNESS: single
:       TOPOLOGY: linear
:   US-09-265-315-78
:
: Query Match      2.0%: Score 33.4; DB 4; Length 2590;
: Best Local Similarity 52.6%: Pred. No. 13;
: Matches 70; Conservative 1; Mismatches 62; Indels 0; Gaps 0;
:
: QY      835  AAAAGGGGAAAAAGGTGCTCTAAACCTAAATGAATGAAGAAGATGCTTCCTATCTTG 894
:         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db      1721  AAAATTGAAGAAGGAAAGATTTGAATTAGATATATATATAAATACGGCAATTTATGCGTTA 1790
:
: QY      895  ATAGATCTCCTTATATATCTTTTATTAACCTCTGCTTGATATGCGCATATGATTTGCATTTA 954
:         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db      1791  TAAAGTATTTTAACTAGTATTTTCAAAAGTATTTGCTTTGCATTCATATATGTCATAT 1850
:
: QY      955  ACTTAATCGAAGT 967
:         ||| ||| |||
: Db      1851  TGTTAAGCTATGCT 1863

```

US-09-266-417-78
: Sequence 78, Application US/09266417
: Patent No. 6228588
: GENERAL INFORMATION:
: APPLICANT: Benton, Bret
: APPLICANT: Lee, Vling J.
: APPLICANT: Malouin, Francois
: APPLICANT: Martin, Patrick K.
: APPLICANT: Schmid, Molly B.
: APPLICANT: Sun, Dongxu
: TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
: TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
: TITLE OF INVENTION: TARGET GENES
: NUMBER OF SEQUENCES: 111
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Word Perfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/266,417
: FILING DATE: March 9, 1999
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/714,918
: FILING DATE: September 13, 1996
: APPLICATION NUMBER: 60/009,102
: FILING DATE: December 22, 1995
: APPLICATION NUMBER: 60/003,798
: FILING DATE: September 15, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 240/248
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 488-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 78:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2590 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-266-417-78

Query Match 2.0%; Score 33.4; DB 4; Length 2590;
Best Local Similarity 52.6%; Pred. No. 13;
Matches 70; Conservative 1; Mismatches 62; Indels 0; Gaps 0;

Qy 835 AAAAGGGGAAAAAGCTTAACCTAAATGATGAAAGATAGCTTCCTATCTTG 894
Db 1731 AAAATTGAAGAGAGATTTGAATTAATTAATAAATAGGCAATTATTCGTTA 1790
Qy 895 ATAGATCTCTTATTTCTTTTATTAATCTCTTGTTATTCCTATGGATTCGATTA 954
Db 1791 TAAGTATTTTTCATAGTTCCTTCAAGATATGTCTTTCATCTCATATTTGCTAAT 1850
Qy 955 ACTTATCGAAGT 967
Db 1851 TGTAAAGCTATGT 1863

RESULT 19

US-09-463-238-2/c
: Sequence 2, Application US/09463238
: Patent No. 6469230
: GENERAL INFORMATION:
: APPLICANT: Edwards, Elizabeth A
: APPLICANT: Smith, Alison M
: APPLICANT: Bustos Guillen, Regla
: APPLICANT: Martin, Catherine R
: APPLICANT: Plant Bioscience Limited
: TITLE OF INVENTION: Starch Debranching Enzymes
: FILE REFERENCE: 97.118
: CURRENT APPLICATION NUMBER: US/09/463,238
: CURRENT FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: PCT/GB98/02280
: PRIOR FILING DATE: 1998-07-30
: PRIOR APPLICATION NUMBER: GB 9716185.5
: PRIOR FILING DATE: 1997-07-31
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 2706
: TYPE: DNA
: ORGANISM: Solanum tuberosum
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (2641)
: OTHER INFORMATION: n = a or g or c or t
: US-09-463-238-2

Query Match 2.0%; Score 33.4; DB 4; Length 2706;
Best Local Similarity 53.0%; Pred. No. 13;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 1460 TTATCTGCAGTACTCTCTTTTCATCATGCTATTTGGTTGTCGAACCTAAGTT 1519
Db 2706 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2647
Qy 1520 AAACAACATATCTTGAAGCGATCTGCTCTTAAGAACAGAGGCTCAAGAAGATTC 1579
Db 2646 CACTANTATTTTATTTTATTAATAATGCAATTAATAACAGAGGTATCTTATAAATCATC 2587
Qy 1580 AGCTCTGCTCTTC 1591
Db 2586 AGCTTCTTCTTC 2575

RESULT 20
US-09-463-238-11/c
: Sequence 11, Application US/09463238
: Patent No. 6469230
: GENERAL INFORMATION:
: APPLICANT: Edwards, Elizabeth A
: APPLICANT: Smith, Alison M
: APPLICANT: Bustos Guillen, Regla
: APPLICANT: Martin, Catherine R
: APPLICANT: Plant Bioscience Limited
: TITLE OF INVENTION: Starch Debranching Enzymes
: FILE REFERENCE: 97.118
: CURRENT APPLICATION NUMBER: US/09/463,238
: CURRENT FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: PCT/GB98/02280
: PRIOR FILING DATE: 1998-07-30
: PRIOR APPLICATION NUMBER: GB 9716185.5
: PRIOR FILING DATE: 1997-07-31
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 11
: LENGTH: 2806
: TYPE: DNA
: ORGANISM: Solanum tuberosum
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (822, 826, 2707 and 2797)

Query Match	2.0%	Score 33.4	DB 1	Length 17410
Best Local Similarity	62.7%	Pred. No. 29		
Matches 52; Conservative	0	Mismatches 31	Indels 0	Gaps 0
3	AAAAAAAACTACGATAGCAAAAT	AAAAAAAAAGTATTCAGAGGCTAAAT	GTGCAAAAACC	62
4401	AAAAAAAAAAGAAAGAAACGAAACG	AAAAAAAAAAGAAACGAAACG	AAAAAAAAAAGAAACG	4342

```

0Y      63 GAAGAAAACCTTTGAAATTT 85
        ||||| | |||||
Db      4341 AAAGAAAGAAAACAAAGAAATTT 4319

RESULT 23
US-08-147-023-3/c
Sequence 3, Application US/08147023
Patent No. 546845
GENERAL INFORMATION:
APPLICANT: OPPERMAN, HERMANN
APPLICANT: OZKAYNAK, ENGIN
APPLICANT: KUBERASAMPATH, THIANGVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H.L.
TITLE OF INVENTION: OSTEOGENIC DEVICES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESS: TESTA, HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/147,023
FILING DATE: 21-FEB-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 810,560
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 827,052
FILING DATE: 28-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 660,162
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,988
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,849
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 616,374
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 600,024
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 599,543
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 579,865
FILING DATE: 07-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 569,920
FILING DATE: 20-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 483,913
FILING DATE: 22-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 422,613
FILING DATE: 17-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 315,342
FILING DATE: 23-FEB-1989
PRIOR APPLICATION DATA:

```


APPLICATION NUMBER: US 579, 865
FILING DATE: 07-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 569, 920
FILING DATE: 20-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 483, 913
FILING DATE: 22-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 422, 613
FILING DATE: 17-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 315, 342
FILING DATE: 23-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 232, 630
FILING DATE: 15-AUG-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 179, 460
FILING DATE: 08-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R.
REGISTRATION NUMBER: 27, 829
REFERENCE/DOCKET NUMBER: CRP-001CP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
FAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17410 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: 3192..3730
OTHER INFORMATION: /label= EXON-1
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 10413..10414
OTHER INFORMATION: /label= GAP-1
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 10696..10891
OTHER INFORMATION: /label= EXON-2
FEATURE:
NAME/KEY: misc-feature
LOCATION: 10960..10961
OTHER INFORMATION: /label= GAP-2
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11059..11211
OTHER INFORMATION: /label= EXON-3
FEATURE:
NAME/KEY: misc-feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11420..11617

OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc-feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
OTHER INFORMATION: THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5
FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: /label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: /label= EXON-7
US-08-447-570-3
Query Match 2.0%; Score 33.4; DB 1; Length 17410;
Best Local Similarity 62.7%; Pred. No. 29;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 3 AATATAAACTCTCAGATAGAAATTAAGTATTCAGAGGCTAATATGCAAAAACC 62
DB 4401 AAAAAAAAAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTAAAC 4342
QY 63 GAAGAAACCTTTGGAAATTT 85
DB 4341 AAGAAAGAAACAAAGAAATTT 4319
RESULT 25
US-08-449-700-3/C
Sequence 3, Application US/08449700
Patent No. 5863758
GENERAL INFORMATION:
APPLICANT: OPPERMAN, HERMANN
APPLICANT: OZKAYNAK, ENGIN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H. L.
TITLE OF INVENTION: OSTEOGENIC DEVICES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,700
FILING DATE: 21-FEB-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 810, 560
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 827, 052
FILING DATE: 28-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 660, 162
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:

```

? APPLICATION NUMBER: US 621,988
? FILING DATE: 04-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 621,849
? FILING DATE: 04-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 616,374
? FILING DATE: 21-NOV-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 600,024
? FILING DATE: 18-OCT-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 599,543
? FILING DATE: 18-OCT-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 579,865
? FILING DATE: 07-SEP-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 569,920
? FILING DATE: 20-AUG-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 483,913
? FILING DATE: 22-FEB-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 422,613
? FILING DATE: 17-OCT-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 315,342
? FILING DATE: 23-FEB-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 232,630
? FILING DATE: 15-AUG-1988
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 179,460
? FILING DATE: 08-APR-1988
? ATTORNEY/AGENT INFORMATION:
? NAME: PITCHER, EDMUND R.
? REGISTRATION NUMBER: 27,829
? REFERENCE/DOCKET NUMBER: CRP-001CP6
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617/248-7000
? TELEFAX: 617/248-7100
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 17410 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: homo sapiens
? FEATURE:
? NAME/KEY: exon
? LOCATION: 3192..3730
? OTHER INFORMATION: //label= EXON-1
? OTHER INFORMATION: //note= "START CODON BEGINS AT POSITION 3313"
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 10413..10414
? OTHER INFORMATION: //label= GAP-1
? OTHER INFORMATION: //note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
? OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
? OTHER INFORMATION: THIS SEQUENCE."
? FEATURE:
? NAME/KEY: exon
? LOCATION: 10696..10891
? OTHER INFORMATION: //label= EXON-2
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 10960..10961
? OTHER INFORMATION: //label= GAP-2
? OTHER INFORMATION: //note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
? OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN

```

```

? OTHER INFORMATION: THIS SEQUENCE."
? FEATURE:
? NAME/KEY: exon
? LOCATION: 11059..11211
? OTHER INFORMATION: //label= EXON-3
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 11351..11352
? OTHER INFORMATION: //label= GAP-3
? OTHER INFORMATION: //note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
? OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
? OTHER INFORMATION: THIS SEQUENCE."
? FEATURE:
? NAME/KEY: exon
? LOCATION: 11420..11617
? OTHER INFORMATION: //label= EXON-4
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 11721..11722
? OTHER INFORMATION: //label= GAP-4
? OTHER INFORMATION: //note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
? OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
? OTHER INFORMATION: THIS SEQUENCE."
? FEATURE:
? NAME/KEY: exon
? LOCATION: 13354..13436
? OTHER INFORMATION: //label= EXON-5
? FEATURE:
? NAME/KEY: exon
? LOCATION: 15044..15160
? OTHER INFORMATION: //label= EXON-6
? FEATURE:
? NAME/KEY: exon
? LOCATION: 17245..17410
? OTHER INFORMATION: //label= EXON-7
? US-08-449-700-3
?
? Query Match 2.0%; Score 33.4; DB 2; Length 17410;
? Best Local Similarity 62.7%; Pred. No. 29;
? Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
?
? Oy 3 AATAAAACATCTAGATACAAATTAAGTATTTTCAGAGCGTAAATATATACAAAACC 62
? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
? Db 4401 AAAAAAAAAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATTAAC 4342
?
? Oy 63 GAAGAAAAACCTTTTGGAAATT 85
? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
? Db 4341 AAACAAGAAAAACAAGAAATT 4319
?
? RESULT 26
? US-08-449-699A-3/C
? Sequence 3, Application US/08449699A
? Patent No. 5958441
? GENERAL INFORMATION:
? APPLICANT: OPPERMAN, HERMANN
? APPLICANT: OZKAYNAK, ENGIN
? APPLICANT: KUBERASAMPATH, THANGAVEL
? APPLICANT: RUEGER, DAVID C.
? APPLICANT: PANG, ROY H.L.
? TITLE OF INVENTION: ANTIBODIES TO OSTROGENIC PROTEINS
? NUMBER OF SEQUENCES: 33
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: TESTA, HURWITZ & THIBEAULT
? STREET: 125 HIGH STREET
? CITY: BOSTON
? STATE: MASSACHUSETTS
? COUNTRY: U.S.A.
? ZIP: 02110
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: /label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: /label= EXON-7
US-08-449-699A-3

Query Match      2.0% ; Score 33.4; DB 2; Length 17410;
Best Local Similarity 62.7%; Pred. No. 29;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 3 AATAAAACTATCAGAAATGAAATAATTTTCAGAGGTAAATATGCACAAAACC 62
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4401 AAAAAAAAAAAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAATTAAC 4342

QY 63 GAAGAAAAACCTTGGCAAAATT 85
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4341 AAAGAAAAGAAAACAAGAAAATT 4319

RESULT 27
US-08-486-343A-1/c
Sequence 1, Application US/08486343A
Patent No. 6071695
GENERAL INFORMATION:
APPLICANT: OZKAYNAK, ENGIN
APPICANT: OPPERMAN, HERMANN
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING
TITLE OF INVENTION: MORPHOGENIC PROTEIN EXPRESSION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
ADDRESSEE: INC.
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 07148

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,343A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, Edmund R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-091CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)-248-7000
TELEFAX: (617)-248-7100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 17415 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: homo sapiens
US-08-486-343A-1

Query Match      2.0% ; Score 33.4; DB 3; Length 17415;
Best Local Similarity 62.7%; Pred. No. 29;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

3 AATAAAACTATCAGAAATGAAATAATTTTCAGAGGTAAATATGCACAAAACC 62

```



```
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 303-3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,494A
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.029
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3989 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 193..1158
US-08-327-494A-1

Query Match          2.0%; Score 33.2; DB 1; Length 3989;
Best Local Similarity 53.0%; Pred. No. 18;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 190 CAAAGACACTCTTATTGGGAGCTCCGCTTGTCGTCAGAGCAATACCTTCATCA 249
    || ||||| || || || || ||||| || || || || || || || || || ||
DB 2938 CAGGACACACTTGTACTGTGCTCAACCGTTCAAGTAAACAGTGTATGTCATGA 2997
QY 250 AGTTTGGCTGTGTCGCCCTGCTATATATCTTATGCTTATTTATGCAAGCTAAGTA 309
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2998 CTCCTAACGCTATGGGCAATTTGCTATATATTTCTGCTATATGTCAGCAAGACA 3057
QY 310 ATATTTTAAGTAAG 323
    | | | | | | | | | |
DB 3058 AAGCCTCCTATG 3071

RESULT 31
US-08-327-494A-3
Sequence 3, Application US/08327494A
Patent No. 5527678
GENERAL INFORMATION:
APPLICANT: Blaser, Martin J.
APPLICANT: Tumuru, Murali K.R.
APPLICANT: Sharma, Smita A
TITLE OF INVENTION: cagB and cagC Genes for H. pylori and
NUMBER OF INVENTION: Related Methods and Compositions
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 303-3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,494A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.029
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3989 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1170..3830
US-08-327-494A-3

Query Match          2.0%; Score 33.2; DB 1; Length 3989;
Best Local Similarity 53.0%; Pred. No. 18;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 190 CAAAGACACTCTTATTGGGAGCTCCGCTTGTCGTCAGAGCAATACCTTCATCA 249
    || ||||| || || || || ||||| || || || || || || || || || ||
DB 2938 CAGGACACACTTGTACTGTGCTCAACCGTTCAAGTAAACAGTGTATGTCATGA 2997
QY 250 AGTTTGGCTGTGTCGCCCTGCTATATATCTTATGCTTATTTATGCAAGCTAAGTA 309
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2998 CTCCTAACGCTATGGGCAATTTGCTATATATTTCTGCTATATGTCAGCAAGACA 3057
QY 310 ATATTTTAAGTAAG 323
    | | | | | | | | | |
DB 3058 AAGCCTCCTATG 3071

RESULT 32
PCT-US95-13659-1
Sequence 1, Application PC/TUS9513659
GENERAL INFORMATION:
APPLICANT: Blaser, Martin J.
APPLICANT: Tumuru, Murali K.R.
APPLICANT: Sharma, Smita A
TITLE OF INVENTION: cagB and cagC Genes for H. pylori and
NUMBER OF INVENTION: Related Methods and Compositions
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 303-3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13659
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.029
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
```



```

:       TOPOLOGY: linear
:       MOLECULE TYPE: DNA (genomic)
US-08-928-361B-3

Query Match                               2.0%; Score 33.2; DB 3; Length 5318;
Best Local Similarity 47.6%; Pred. No. 20;
Matches 98; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

OY 1053 GFGATACCGCTACTTATTCATGCTATTTGTTGGGATACGTCATTCGTAAATTTGGATGG 1112
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 1061 GTAGTACTACTTGTGTTGTTGTTGGTAGTACTAGTACTAGTACTGTTGTTGTTGTTGTTG 1002
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
OY 1113 TTAACGTGAGCCCTAGTCACTCCGTGATGTTCTCCTAACAGATACGTTTCTTCGCT 1172
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 1001 GFGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 942
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
OY 1173 CTGTATCTTTTGAACACAGCTTCTGGCTGCTGCTATGTTGCGTACACTCTCTC 1232
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 941 GTTGTGTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 882
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
OY 1233 ATGCTAGCTGTGTTGTTGCGAGCTAT 1258
      |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 881 GTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 856

RESULT 38
US-09-020-956-140
: Sequence 140, Application US/09020956
: Patent No. 6261562
:
: GENERAL INFORMATION:
:
: APPLICANT: XU, Jiangchun
: APPLICANT: DILLIN, Davin C.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FOR
: NUMBER OF SEQUENCES: 178
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentia Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/020,956
: FILING DATE: 09-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: MAKI, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO.: 140:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 200 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
:
: US-09-020-956-140

Query Match                               2.0%; Score 33; DB 4; Length 200;
Best Local Similarity 54.1%; Pred. No. 5.8;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Y 150 TTCCTTCGTATTACATTAACTACGAGTGTACCGCATACAAAGACACCTGTATTGTG 209

```

[illegible]

```

; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Ketter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 140
; LENGTH: 200
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(200)
; OTHER INFORMATION: n = A,T,C or G
US-09-605-785-140
```

```

Query Match          2.0%; Score 33; DB 4; Length 200;
Best Local Similarity 54.1%; Pred. No. 5.8;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 150 TTCTTGTGATTAACATTAACTATACGGTGTAGCGGATACAAAGACACTTATTGTG 209
      |||| ||| ||| |||| |||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13 TTCTGTTGTGTGATTTAGTATAGGGGTTNGCTTNTCTAANATACTTTCAATTA 72

QY 210 GGAGCTCCCTGGTCTGTGTCAGAGGCAATACCTTTCATCAAGTTTGGCTT 260
      || || || || || || || || || || || || || || || || || ||
DB 73 ACANCTTTTGTTAAGTGTGAGGCTGCACCTTGCTCCATANAATTAATTGTTT 123
```

Search completed: January 21, 2003, 19:33:00
Job time : 354 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 18:00:23 : Search time 1674 Seconds
(without alignments)
15837.522 Million cell updates/sec

Title: US-09-869-433-1
Perfect score: 1637
Sequence: 1 gaataaaactacacagaa.....gtgatccctactcgtcttt 1637

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	332.2	20.3	1866	11 AY107360	AY107360 Zea mays
2	226.4	13.8	966	10 BE574942	BE574942 D06 Triph
3	197	12.0	928	14 H55010	H55010 HH042a Sorg
4	195.8	12.0	600	12 BG097617	BG097617 EST462136
5	190.4	11.6	567	10 AW931058	AW931058 EST356901
6	184.4	11.3	691	12 BG269363	BG269363 L0-3415T3

7	181.4	11.1	558	14 B0790720	B0790720 E4555 Chi
8	178	10.9	645	10 AV926113	AV926113 AV926113
9	168.4	10.3	664	13 B1405734	B1405734 115G06 Ma
10	168.2	10.3	715	10 BE459811	BE459811 EST415103
11	165.6	10.1	568	12 BG463286	BG463286 EML_48_F1
12	165.2	10.1	602	14 B0754402	B0754402 EBCa01_SQ
13	165.2	10.1	563	13 BM448586	BM448586 DSA025604
14	163.8	10.0	685	14 B0490006	B0490006 26-E9427-
15	163.6	10.0	573	13 B1300874	B1300874 B1300874
16	163.2	10.0	527	9 A499428	A499428 A499428
17	163.2	10.0	529	13 B1271311	B1271311 NF051C03F
18	162.6	9.9	678	9 A432835	A432835 A432835
19	160.8	9.8	787	13 B1177811	B1177811 EST518756
20	158.4	9.7	518	10 AW738613	AW738613 EST340040
21	158	9.7	686	14 BQ255402	BQ255402 MTNAC62TK
22	157.8	9.6	637	13 B482169	B482169 B482169
23	157.4	9.6	624	12 BF460154	BF460154 066G08 Ma
24	155.8	9.5	567	14 BM954156	BM954156 sam70h06
25	155.6	9.5	609	13 BM322252	BM322252 PIC1_2_B0
26	154	9.4	532	14 BQ469843	BQ469843 HX01A14T
27	149.6	9.1	700	14 BQ147136	BQ147136 NF034F08F
28	148	9.0	499	12 BF459810	BF459810 065F05 Ma
29	148	9.0	567	14 BQ507193	BQ507193 EST614608
30	147.2	9.0	462	9 AU057209	AU057209 AU057209
31	146.8	9.0	529	13 BM357837	BM357837 1211T-F7
32	146.2	8.9	469	14 C74932	C74932 C74932 Rice
33	146	8.9	581	10 AN924705	AN924705 WS1_71.D0
34	144.6	8.8	522	12 BG322472	BG322472 EML_28.D0
35	143.8	8.8	744	12 BF460035	BF460035 072A12 Ma
36	143.6	8.8	602	12 BF053092	BF053092 EST438322
37	143.6	8.8	667	10 AM509457	AM509457 s137d04_37
38	143	8.7	606	12 BF423978	BF423978 s502f04_Y
39	142.8	8.7	663	12 BG456211	BG456211 NF075B07F
40	141	8.6	575	12 BE726529	BE726529 894092F04
41	140.8	8.6	557	10 AM671962	AM671962 LG1_353_G
42	140.8	8.6	648	10 AM255797	AM255797 ML866 pep
43	138.8	8.5	601	10 AM255224	AM255224 ML219 pep
44	136.4	8.3	556	13 B1463735	B1463735 B1463735
45	136.2	8.3	477	9 AJ499380	AJ499380 AJ499380

ALIGNMENTS

RESULT 1	AY107360	1866 bp	mRNA	linear	HTC 25-MAY-2002
LOCUS	AY107360				
DEFINITION	Zea mays PC0098455 mRNA sequence.				
ACCESSION	AY107360				
VERSION	AY107360.1	GI:21210438			
KEYWORDS	HTC.				
SOURCE	Zea mays.				
ORGANISM	Zea mays				
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: PACC clade: Panicoideae: Andropogoneae: Zea.				
AUTHORS	1 (bases 1 to 1866) Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.				
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2 (bases 1 to 1866) Coe,E.C.				
AUTHORS	Direct Submission				
TITLE	Submitted (25-Apr-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
JOURNAL	Location/Qualifiers				
FEATURES	1..1866				
SOURCE	/organism="Zea mays" /db_xref="MaizeDB:635637" /db_xref="taxon:4577" /clone="PC0098455"				

BASE COUNT ORIGIN	437 a	446 c	467 g	516 t
/clone_11b="Maize Mapping Project/Dupont Consensus library" /note="this sequence is part of a project of ESR assemblies resulting from the application of public configs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"				

	Query Match	20.3%	Score 332.2	DB 11	Length 1866:
	Best Local Similarity	55.0%	Pred. No. 3, 3e-67:		
	Matches 725:	Conservative 0;	Mismatches 578;	Indels 16;	Gaps
OY	226 GTGAGAGGCATACCTTTTCATCACAAGTTTTGGCTTGTTGTCCCTGTGCTATTAATCTTAA	285			
Db	18 GCGCGAGATCATCCCTTCCTCAAGCCGTGGTCAACCGGCCATGGCATCGGGCTCA	77			
OY	286 TGCATTATTTATGCAAACTCAAGTAATAATTTAATGAAGGCGCTAATTTTATGAAGCTG	345			
Db	78 TGCTCTCTACACCACACTGCGGATGCTCTCTCCAAAGGAGGGCTCTTCTACGGCGTCA	137			
OY	346 GAAGCCCCCTTTTAAATTTCTTTGGCCCTGTCTCCGACTGTAATTTATPCGACAGGATG	405			
Db	138 TCCTTCCGGTTCATCGCCTCTCTGGGGGCGCTTGGCCTACAGTGTCTACCCATGCGGAG	197			
OY	406 TTTTACATCTACAGAATTTTGCTGACCGCTTTACAGGCATCTCACTCCAGATTTGCTAG	465			
Db	198 CCATCTACACCCACACGCGCTCGCCGATCGCTCTCGCTCGCGGCGCCACCTTCCTG	257			
OY	466 GACTCGTTGGCATCTTTAAGAAACGGACATTTCTCGATTTTATGTACTGTGTAACAT	525			
Db	258 GACCCGTTGGCATCTCTCCGGCTGTGAGCTTCTGCTCTTTACGTCATGCGCGAGCTT	317			
OY	526 GGGAAGACGTCATGCTACTCTATGTCTTGAGGGAATTTGCTAATGAATTAACAAAATCC	585			
Db	318 GGGGCACTGTCGTCATATCCGTCCTTTTCTTGSGGGTTTGCAATCAGATTACTACAGTTG	377			
OY	586 ACGAAGCAAAGCGTTTCTACGCTCTTTTGGGTATGCGAGCTAATATTCTTTACTAGCTT	645			
Db	378 AAGGGGCCAAAGAGTTCTTACCCCTGTTTGGGCGTATGGGCCAATGGGCCCATCTTCT	437			
OY	646 CTGCTCGTCGAATTTGTTGGCGTTCCAAGTTGAGACCTTCGTTCTGAAGGTGATGATC	705			
Db	438 CTGGGCGCACGGTGAATATTCTCAACACTGAGAAAGATTTGGGTCCTGGAGTGGATG	497			
OY	706 CTTGGGGAATTC-TTTTACGCTTTTGATGCGCTATGACTATTTGTATCTGGACTTGTCTT	764			
Db	498 GTTGGGCAATTTCTGTTGAAGGACATGATGAGCATAGTGGTGTACTGGCTTGTGCATC	557			
OY	765 ATGSCCAGTTACAGCGGTGATCANTAAAGCGTTTACGCGATCTCCGCTTATTAATACCA	824			
Db	558 ACCGGTATCTATTATGGGAGTAGAACAATTTGTTATTGATTAATC-----ATCT	605			
OY	825 GAACAAATGCAAAAGGGGAAAAAGGTGCTAAACCTAAATGAATATGAAGAATAGCTTC	884			
Db	606 GTGCCGCTGGTTGAGGGGAAAAAGAGACCAACCACTCAGCATGGGTGAGAGCATG	665			
OY	885 CTCATCTTGATAGATCTCCTTATATTCTTTTATFACTCTCTTGTTATTTGCTAATGTT	944			
Db	666 AAGGTTCTGTTGATCTCGATATGTAGAGGATCTTGCTACATATGTTGTTCTTATGGA	725			
OY	945 ATTGGCAATTAACATAACGAGTGTGGAAAGACACTGAACTGAACTGCAATATCTAAT	1004			
Db	726 ATAAGCAATTAACCTTGTCAGAGTGACATGSAATTCAAAATTTGAAGCACAGTTCCCAAC	785			
OY	1005 ATGAATGACTATAGTAGTTCATGGGAACTTCTCTTGAGACTGGCGCTAGTATCCGTA	1064			
Db	786 CCGAATGAATATTCTTCATTCATGGGGAATTTCTCAACGTCCACGCGGATAG---CTACA	842			
OY	1065 CTATATCATCTATTTTGTGGTGGTAACGTCAITTCGTAAATTTTGGATGGTTAACTGAGCC	1124			

Db	843	TTTACAAATATGTTGTTAGGAAGAGTATCCTCAAGAAAGTTGGGTGGGAGATTGCAGCT	902
Qy	1125	CTAGTCACCTCTGTCTATGGTCTTCTCTTAACAGGTATCGTTTCTTGCTCTTGTATCTTT	1184
Db	903	ACGATCAGCCCTGCAGTGTATTACCTCCACAGGAGTTGGGTCTCTCACTGATTTTGTGTT	962
Qy	1185	ACGAACCAAGCTTCTGGGCTGGTCCGTATGTCGTAGTTGGTACAACTCTCTCATGTCAAGCTGTG	1244
Db	963	GGTAGCCATTTGATCCTCTTATGACCAAGTTGGATGACACCTTTCCTGGCGCAGTC	1022
Qy	1245	GTTTGCGGAGCTATACAGATATCTTTCGAAATCCACAAGTTAGCTCTTTTGACTTA	1304
Db	1023	TATGTGGAGCAATGCAGAACATTTTTCAGTAGAGTGCAGAAATACAGTCTGTTTGTATCT	1082
Qy	1305	ACTAAGAATAGGCTATATCCCTCTTGACCAAGACCAAAAGTCAAAGSTAAAGGTGCT	1364
Db	1083	TGCAGGAAGATGGCATACATCTCTTTGGATGAGGATATGAAGCTGAAGAGTAAAGCGGCT	1142
Qy	1365	ATTGATGTATGGCCGCCCGCTTGGAAATACAGGAGAGCTTTATACACAAAGSTTTTG	1424
Db	1143	ATTGATGTTGTGTGCACACCCCTTTGGGGAATCTGGAGGTGCTGTGATCCAGCAGTTTCATG	1202
Qy	1425	CTCGTATCTGTGGAAGTATTGGAGCTATGAGACCCCTTATCTTGCAATGATTTCTCTTTTC	1484
Db	1203	ATTCGTCTATTGGTTCCTCTCGGAACTGCACACCTTACCTTGGGCGAATTACTCCTGGTG	1262
Qy	1485	ATCATTTGCTATTGGTGGTTCTGCACAACTAAGTTAAACAATATTCTTACGCGAGTC	1543
Db	1263	ATTGTCTTGCATGGCTGGGTGCTGTAAGGTCCCTGCACCTCAGTTCCTTCCCTGGC	1321

```

RESULT 2
LOCUS      BE574942
DEFINITION BE574942 966 bp mRNA linear EST 07-DEC-2001
            DDB Triphysaria versicolor root-tip, early DMB0-induced transcript
            cDNA library Triphysaria versicolor cDNA, mRNA sequence.
ACCESSION  BE574942
VERSION     BE574942.1 GI:12001272
KEYWORDS    EST.
SOURCE      Triphysaria versicolor.
            Triphysaria versicolor.
            Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Lamiales; Orobanchaceae; Triphysaria.
            1 (bases 1 to 966)
REFERENCE   Matvienko, M., Torres, M. J. and Yoder, J. I.
            Transcriptional responses to the hemiparasitic plant Triphysaria
            versicolor to host plant signals
            Plant Physiol. 127 (1), 272-282 (2001)
21437952
COMMENT     Contact: Yoder, J. I.
            John I. Yoder Research Lab, Dept. of Vegetable Crops
            University of California at Davis
            137 Assmannson Hall, One Shields Drive, Davis, CA 95616, USA
            Tel: 530 752 1741
            Fax: 530 752 9659
            Email: jiyoder@ucdavis.edu
            length = 966 bp.
FEATURES
            Source
            Location/Qualifiers
                1..966
                /organism="Triphysaria versicolor"
                /db_xref="taxon:64093"
                /clone_lib="Triphysaria versicolor root-tip, early
                DMB0-induced transcript cDNA library"
                /tissue_type="root-tips ~5mm length"
                /dev_stage="3-4 weeks growth"
                /lab_host="E. coli"
                /note="Vector: pCR2.1 TA Cloning System, Invitrogen,
                Carlsbad, CA; PCR-based suppression subtractive
                hybridization cDNA library"
222 a 186 c 240 g 318 t
ORIGIN

```

Query Match 13.8%; Score 226.4; DB 10; Length 966;
 Best Local Similarity 53.7%; Pred. No. 1.8e-42;
 Matches 532; Conservative 0; Mismatches 431; Indels 27; Gaps 2;

QY 298 CAAAGCTAAGTAATTTTAAAGTAAAGCCCTTAATTTTATGACAGTGGGAACCCCTTTT 357
 DB 4 CTAAGTTGGCTAATGTTTGTGGAAGATGACATTTCTACACTCTTATGTGGCATTTCA 63
 OY 358 TAATTTTCCCTGCGCTCCGACGTAATTTATCCGCTACGCGATTTTATACATCTCA 417
 DB 64 TTGCGTTTGTGGGACATTTGGGCTTTGCTTATCTCTCAGTAAATTTTTCACCCCA 123
 OY 418 CAGAATTTGCTGACCGCTTACAGCCATCTACCTCAAGATTTGCTAGACATCTGTGCA 477
 DB 124 CTGGCGCTTGTGATTAAGCTTTTGAACGTTGGGCTCCAGGTTCTTGAGACACTGTGCA 183
 OY 478 TCTTAAGAACTGACATTTGCTGATTTTATGATCTTGTGACATTAAGGGAAGCTCA 537
 DB 184 TTTTGGAGATTTGGAGCTTTTGTGTTCTATGATGCTGACGCTTTGGGCGACGTGCG 243
 OY 538 TGCTATCTATGCTTCTGCGGATTTGCTAATGAATTAACAATAATCCAGCAAGCAAGC 597
 DB 244 TTGTTCTGTTTTGTTTGGGGGTTTGAATCAGATTAACACACATGACGACGCAAGA 303
 OY 598 GTTCTACGCTCTTTCCGCTACGAGCTAATTTCTTACTAGCTTCTGCTGCA 657
 DB 304 AATTCATACCTTTTGGGCTCCGAGCAAAATGACATCTTCTCGGCTGGACTG 363
 OY 658 TTTGTTGGGCTTCAAGTTGAGAGCTTCCGTTCTGAAGGTGATGATCCTTGGGAATTT 717
 DB 364 TGAATATCTTTCCCAATGAGGCAAAATTTGGCCCGGCTGATGTTGGGCACT 423
 OY 718 CTTTACGCTTTTGTGCGCTATGATATGATCTGACATTTCTTATGGCACTACT 777
 DB 424 CTTTGAAGGAGATGATGATTTGATGAGGGAGGCGCTTATGATTTGTTCTATAT 483
 OY 778 GGTGATCAATAGAACCTATTTGACCGATCTCGCTTATATATCCAGAAAGATGCAAA 837
 DB 484 GGTGGTAAACATTAATGT-----TCCTCTCGGACCGTA 519
 OY 838 AGGGGAAAAAGGTCTTAACCTAATGAATGAATGAAGATAGCTTCTATCTTGTATA 897
 DB 520 GTATGAAAGAGAGAGAGAGCCGAAATGAGACATGAGAGCTTAAAGTTCTGTAT 579
 OY 898 GATCCCTATATCTTTTAACTCTCTGTTATGCGATGCGATGCTTTGCACTTACT 957
 DB 580 CAAAGAAATATATTAAGATTTGGCCACTTTGTTGTCATATGCAATTAACCC 639
 OY 958 TAATCGAAGTACTTGAAGAAGTCAAGTGAACCTGCAATATCTAATATGAATGACTATA 1017
 DB 640 TGTGTCAGGTTACATGAAATCAAAAGCTCAAGTCTCCCAAGCGGCAAGCAATAT 699
 OY 1018 GTGAGTTCATGGGACCTTCTCTCTGACATGCGGTGATATCGTACTTATCATGCTAT 1077
 DB 700 CTGCTTCATGAGGATCTTTTCACTACTACCGAATAGCAAACTTCAATATGATGCTTT 759
 OY 1078 TTGTTGGTGAAGCTCAATGTAATTTGATGTTGATGCTTAACTGAGACCCCTAGCATCTG 1137
 DB 760 TGAGTCAATGATTTTAAAC---AAATAGCGTTGGGATGACACCAAGATTAAGCGCA 816
 OY 1138 TCATGGTCTCTCTACAGATATCTGTTTCTCTGCTTGTATCTTAAAGCAAGCTT 1197
 DB 817 CCGTCTCTATACCTCTGGGATTTGCTTCTCTGTTCTTCTGTTGCTGACCTTTTG 876
 OY 1198 CTGGGCTGCTGATGTTGCGTACACTCTCTCATCTAGCTGATGTTGTGCGAGCTA 1257
 DB 877 CCCCCTGCTCGCGCGCTTCGGAATGACTCCTATATAGCAGCTGTATGTTGGGCGCA 936
 OY 1258 TACAGAAATATCTTCGAATCCCAAAAT 1287
 DB 937 TGCAAAACATTTTATGAAGAGCGCAAGT 966

RESULT 3
 H55010
 LOCUS
 DEFINITION HH042a Sorghum bicolor cv. TX430 Sorghum bicolor CDNA clone HH042
 5' similar to adenine nucleotide translocator, chloroplast, mRNA
 sequence.
 ACCESSION H55010
 VERSION H55010.1
 KEYWORDS GI:1000790
 SOURCE
 ORGANISM
 sorghum.
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE
 1 (bases 1 to 928)
 Wyrich, R., Dreesen, U., Brockmann, S., Streubel, M., Chang, C., Qiang
 D., Paterson, A.H. and Westhoff, P.
 The molecular basis of C4 photosynthesis in sorghum: isolation,
 characterization and RFLP mapping of mesophyll- and
 bundle-sheath-specific cDNAs obtained by differential screening
 Plant Mol. Biol. 37 (2), 319-335 (1998)
 JOURNAL
 MEDLINE
 COMMENT
 CONTACT: Westhoff P
 Plant Molecular Biology
 Heinrich-Heine-University
 Universitaetsstrasse 1, D-40225 Dueseldorf, Germany
 Tel: (49) 211-311-2338
 Fax: (49) 211-311-4871
 Email: west@uni-duessel.dorf.de
 Seq primer: SK (17-mer)
 FEATURES
 source
 1..928
 /organism="Sorghum bicolor"
 /strain="cv. TX430"
 /db_xref="taxon:4558"
 /clone="HH042"
 /clone.lib="Sorghum bicolor cv. TX430"
 /note="vector: Lambda ZAP II; leaf cDNA library,
 unidirectionally cloned"
 BASE COUNT 225 a 182 c 245 g 275 t 1 others
 ORIGIN
 Query Match 12.0%; Score 197; DB 14; Length 928;
 Best Local Similarity 54.5%; Pred. No. 1.4e-35;
 Matches 511; Conservative 0; Mismatches 405; Indels 21; Gaps 5;

QY 607 CTCTTTGCGATCGAGCTAATATTTCTTACTAGCTTCTGCTGTCGCAATGTTGGG 666
 DB 1 CGCTGTTGGCGCTTGCCCAATGCGCCCTACATCTCTGCGGCGACGGTGAATAT 60
 OY 667 CTCAAGTTGAGAGCTTCGTTCTGAAGGTGATGATCTTGGGGAATTTCTTACGTC 726
 DB 61 TCTCAACATGAGGAAGATTTGGGCTCCGAGTGAAGTGGGCAATTC----- 112
 OY 727 TTTTATGCGTATGACTATTTGATCTGAGCTTGTCTTATGCGCATGCTTACGATGATCA 786
 DB 113 -GTTAAAGCCATGAGCATATGATGCTTACTGCTTATCATCACCGGTATCTATTGGG 171
 OY 787 ATAGAACGTAATGACCGATCTCGCTTCTATATTCAGAAAGAAATGCAAAAGGGGAAA 846
 DB 172 GGGTCAACAGTTTGTATATGATACATCATGATCC-----GTGGTCAGCGGAAA 224
 OY 847 AAGGTCTAAACCTTAATATGAATGAAGATAGCTTCTCTATCTATGATGATCTCTT 906
 DB 225 AGAAGCAAGCAAGCAAGCTGACGATGGGTGAGAGCATGAAGGTTCTGTTGATCTCCGT 284
 OY 907 ATATCTTTTATTAATCTCTGTTGTTATGCTTATGCTATGATGATTAATTAATGCAAG 966
 DB 285 ATGTAGGAGATCTTCCACATTTGGTCTGTTCTTATGATTAAGCATTAACCTTGTGGAGG 344
 OY 967 TGACTTTGAAAAAGTCAAGTGAATGCAATATCTTAATATGAATGATGATGATGATGATCA 1026
 DB 345 TGACATGCAAAATCAAAATTTGAAGCAGAGT--TCCAAGCCCAAAATGATATCTTATTTCA 403

QY	1027	TCGGGAACCTTCCTTCCTGACGGGGGTGATACCGCTACTTTCATGCAATTTGTTGGG	1086
QY	1087	GTAAGCTCATTCGTAAATTTGGATGGATTACGTGAAGCCCTAGTCACTCCCTGCATCGTTTC	1146
QY	404	TGGGGCATTTTCTCAACTGCCACCGCATAGCTTTCATTAATGATGTTGTTAGGAGAG	463
Db	464	TAAATCTCA---GAAAGTTCCGGGTGGGAGTTGCACTACGATGACATCCGACGTGTTAC	520
QY	1147	TCCCTAACAGATATCGTTTCTTCGCTCTTGTTATCTTTAGAAACCAAGCTTGGGCGTGG	1206
Db	521	TTCTACACAGAGATGGGTTCTTCCTCACTGATTTTGTGGTGAAGCATTTGACCTCCCTTA	580
QY	1207	TGCGTATGTTTCGGTACAACTCCTTCATGCTCTAGCTGTGGTTCGCGATATACGAATA	1266
Db	581	TCACCAAGTTTGGGATGACACCTTTGCTTGGGCGACTCATATTGTSAGCAATGCGAACA	640
QY	1267	TTCTTTGCAATTCACAAATATACGCTCTCTTTGACTCAACTAAGAAATAGCCTATATCC	1326
Db	641	TTTTAGTAAAGAGTCAAAAGTACAGTCTGTTTGATCTTTGCAAAAGAGATGATACATTC	700
QY	1327	CTCTTGACCAACAGCAAAAGTCAAAAGGTAAGCGCTATTTATGATGATGCGCCCGCT	1386
Db	701	CTTTGGATGATGAGATATGAAAGGTTAAAGGTAAGCAGCTATTAGCTGTGCTCAACCCCT	760
QY	1387	TCGGAATATCAGAGAGAGGCTTTATTCACAACAAGTTTGCCTTATTCGTGAGAGATTG	1446
Db	761	T-GGGAATCTGAGAGTGGTTGATTCACAGCACTTCATATCCTGTCACTTTGGATCTCTTG	819
QY	1447	GAGCTATGACCCCTTATCTTGCGATGATTCCTTTTCATCATTCCTATTTGGTGGTTT	1506
Db	820	CGAATTCGACACCCCTACCTTGGTGGAAATCTCTGTGATGTTCTTGATGATGGCTGGTG	879
QY	1507	CTGCAACTAAGTTTAAACAACATATTCTTAGGCCAGTC	1543
Db	880	CTGTAAAGTCCCTCGACATCAGATTTTCAACCCCTGGC	916
RESULT 4	BC097617		
LOCUS	EST1462136	potato leaves and petioles Solanum tuberosum cDNA clone	
DEFINITION	ESTB47J17 5' sequence, mRNA sequence.		
ACCESSION	BC097617		
VERSION	BC097617.1	GI:12587652	
KEYWORDS	EST.		
SOURCE	potato.		
ORGANISM	Solanum tuberosum		
REFERENCE	Eunayota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterisids I; Solanales; Solanaceae; Solanum.		
AUTHORS	van der Hoeven, R.S., Bezzerides, J., Holt, I.E., Liang, F., Cho, J., Uteback, T., Hansen, C.L., Doan, B., Bougri, O., Buell, C.R., Ronning, C.M., Fry, W.E., Tanksey, S.D. and Baker, B.		
TITLE	Generation of ESTs from potato leaves and petioles		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cconatresgen.com. Location/Qualifiers		
FEATURES	source		
	1..600		
	/organism="Solanum tuberosum"		
	/cultivar="Kennebec"		
	/db_xref="taxon:4113"		
	/clone="ESTB47J17"		
	/clone_lib="potato leaves and petioles"		
	/tissue_type="leaflets and petioles"		
	/dev_stage="8 weeks old plants"		
	/lab_host="SOLR"		
	/note="Vector: phuescript SK(-); Site_1: EcoRI; Site_2: XhoI; Tissue was supplied by Dr. Fry (Cornell University)."		

[illegible]

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES

Location/Qualifiers

1. 567
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEF42H20"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; CLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 113 a 95 c 138 g 221 t
ORIGIN

Query Match 11.6%; Score 190.4; DB 10; Length 567;
Best Local Similarity 60.0%; Pred. No. 4.9e-34;
Matches 317; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

OY 212 AGCTCTGCTGCTGTCGACAGGCAATACCTTCATCAAGTTGGCTTTGTCCTCG 271
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 2 AGCTAAAGGCTCTAGCTGACAGATATCTTCTTGAAGAACTGGTGAATTTGCTAT 61
OY 272 TGCTTTTACTTATCTATTTATGCAAGCTAGTATATTTTATAGAACAGCCTT 331
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 62 GCGTATGATTCATCTTTTGTACACAAAGTTGCTATGTTGTCAAGAGAGCTCT 121
OY 332 ATTTATGACAGTGGAGAGCCCTTTTAAATTTCTTGGCCCTGTCGCGACTATTTA 391
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 122 TTTTATGCTTATACCTTTTATTTGATTTGCGGCTTTGGTGTGTTTGA 181
OY 392 TCCGTCAGCGATGTTTATCATCTACAGAAATTTGCTGACCGCTTACAGCCATCT 451
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 182 TCCTCTTACGCAATTTCTTCAACCTTACAGCTTTGCTGTAAGCTTCAATATCCCT 241
OY 452 TCCAGATTTGCTAGACCTGCTGATCTTAAAGAACTGCATTTGCTGCTATTTAT 511
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 242 TCCAGATTTTCTTGACCAATTTGCTATCTGAGATCTGAGATTTCTGCTTTTAT 301
OY 512 ACTGCTGAACATATGGGAGAGCGTATGCTATCTAAATGTTCTGGGATTTGCTAAT 571
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 302 CATGCTGAGCTTTGGGAGAGTGTGTTTCTAGTACTTTTGGGATTTGCTAATCA 361
OY 572 AATTCAAAAATTCACGAAGCAAGCGTTTCTAGCCTCTTTTGGTATCGAGCTAATAT 631
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 362 GATTACGACTGTGATGAGGCTAAGAAATTTCTATCTTTGTTGGAAGCTTGAGCAAT 421
OY 632 TTCTTACAGCTTCTGCGCTGCAATTTGTTGGCTTCAAGTGAAGGCTTCGCTTTC 691
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 422 TGCTTTTATTTTCTTGCGGCAAGTGAAGTATTTCTTCTAGAGGCTCTTTAAG 481
OY 692 TGAAGGTAGATCCCTTGGGGAATTTCTTACGCTTTTGTATGAGCTAT 739
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 482 TCCTGAGATGATGTTGGGCTATCTCCCTGAAGAAATGATGAGCAT 529

RESULT 6

LOCUS

DEFINITION BG269363 691 bp mRNA linear EST 20-FEB-2001
L0-3415T3 Ice plant Lambda Uni-Zap XR expression library, 0 hours
NaCl treatment Mesembryanthemum crystallinum cDNA clone L0-3415 5',
mRNA sequence.

ACCESSION
VERSION

BG269363
BG269363.1 GI:12975326

KEYWORDS

EST.
common iceplant.
SOURCE
ORGANISM

Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.

REFERENCE

1 (bases 1 to 691)
Cushman, J.C.

AUTHORS

An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum

TITLE

Unpublished (1997)

JOURNAL

Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L0-35 row: B column: 3
Seq primer: T3
High quality sequence stop: 350
POLYA-No.

FEATURES

Location/Qualifiers

1. 691
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="L0-3415"
/clone_lib="Ice plant Lambda Uni-Zap XR expression library"
/tissue_type="leaf"
/dev_stage="Six week old"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT

166 a 149 c 173 g 203 t
ORIGIN

Query Match 11.3%; Score 184.4; DB 12; Length 691;
Best Local Similarity 57.5%; Pred. No. 1.2e-32;
Matches 351; Conservative 0; Mismatches 256; Indels 3; Gaps 1;

OY 929 GGTATTTGCTATGATTTGATTTGATTAATTCATATTCAGTATGGAAGCAAGCTGTA 988
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 1 GGTGTTGATATGATATGATTAATTCATATTCAGTATGGAAGCAAGCTGTA 60
OY 989 ACTGCAATATCCTAATATGATGATGATTTGATGAGGAGCACTTCTCTGAGC 1048
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 61 ACCTCAGTTTGGAGGCCCAATGAAATATTTCTGTCATGAGTGATTTCTACGCAAC 120
OY 1049 TGCGCTAGTATCCGACTATATCATGCTATTTTGGTGTGTAAGTCATTTGGA 1108
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 121 TGGGATATGCAACCTTCTCAATATGATGCTTCTGAGCCAGTTTCAATTCGACAAGTT 177
OY 1109 ATGGTTAATGAGCCCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1168
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 178 CTGGGAGTTGCTGCTGCCATCAGCTTACGCTTCTTCTTCTTCTTCTTCTTCTTCT 237
OY 1169 CCGCTTTGTTATTTAGAAACCAAGCTTCTGGGCTGCTGCTATTTGCTGTAACCTC 1228
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 238 TTCAATGATTTTGTTCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
OY 1229 TCTCATGCTAGCTGTGTTGTGAGACTATACAGAAATATTTTGTGAAATCCACAATA 1288
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 298 TCTTCTAGCAGCATATATGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357
OY 1289 GCGTCTTGTGATCAACAAAGAAATGGCTATATCCCTTGTGCAAGAGCAAAAGT 1348
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 358 CAGTTTGTGTTGATCTTCTGCAAGAAATGGCTTATCATTTGATGAGACACTAAGCT 417
OY 1349 CAAAGTAAAGCTGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1408
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

Db	Accession	Source	Organism	Reference Authors Title	Journal Comment	Features
Db	1418	TAAGGGGAAGGACGATTTGACGTGGTTTGCAACCGCTTGGAAATATCCGAGGGGCTCT	477			
Qy	1409	AATCCAAACAGSTTTGCTCTTATCTGTGCGAAGTATGAGAGTATGACCCCTTATCTTGC	1468			
Db	478	GATCCAGCAGTTCATGATCATCTTAACCTTTGGGACACTTGCCAAATTCACCCCTTACCTTGG	537			
Qy	1469	ACTGATTTCTTTTCATCATCTGATTTGGTTGGTTTCTGCAACCAATTAACCAACT	1528			
Db	538	AGGAAATCTCTTGTGATTTGATCACTGCATGCTGGCAGCAGCAAGCTTTTGACAAACA	597			
Qy	1529	ATTCTTAGCG	1538			
Db	598	GTTTACAGAG	607			
RESULT 7						
LOCUS	B0790720	558 bp	mRNA	linear	EST 30-JUL-2002	
DEFINITION	E4555 Chinese cabbage etiolated seedling library	Brassica rapa				
ACCESSION	B0790720					
VERSION	B0790720.1	GI:22005682				
KEYWORDS	EST.					
SOURCE	Brassica rapa subsp. pekinensis.					
ORGANISM	Brassica rapa subsp. pekinensis.					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.					
	Ryu, S.H., Yang, K.A., Lee, S.Y., Kim, H.-I., Cho, M.J. and Lim, C.O.					
	Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA					
	(2002)					
JOURNAL	Unpublished (2002)					
COMMENT	Contact: Lim, C.O.					
	Plant Molecular Biology & Biotechnology Research Centre					
	Gyeongsang National University					
	#900 Gazwa-dong, Jinju 660-701, Korea					
	Tel: 82 55 751 6235					
	Fax: 82 55 759 9363					
	Email: collim@nongae.gsnu.ac.kr					
	Seq primer: T7.					
FEATURES						
source	location/qualifiers					
	1..558					
	/organism="Brassica rapa subsp. pekinensis"					
	/cultivar="Jangwon"					
	/db_xref="taxon:51351"					
	/clone="E4555"					
	/clone_lib="Chinese cabbage etiolated seedling library"					
	/tissue_type="Etiolated seedling"					
	/lab_host="XL-1 Blue"					
	/note="Vector: pSFORM 1; Site_1: Sal I; Site_2: Not I"					
BASE COUNT	131 a 141 c 131 g 155 t					
ORIGIN						
Query Match	11.1%	Score 181.4	DB 14	Length 558		
Best Local Similarity	59.7%	Pred. No. 6.2e-32				
Matches 324	Conservative 0	Mismatches 216	Indels 3	Gaps 1		
Qy	936	GCGTATGATTTGGCATTAACCTTAATGCAAGAGCACTTGGAAAGAGCTGAACTGC	995			
Db	13	GCGTATGATTTAGTATCAACCTTGTGTAAGCATGATGAAATTCAAACCTTAACCTGC	72			
Qy	996	TATCCTATATGAAATGATATAGTAGTTTCATGGGAGCACTTCTCTGACCTGGCGTA	1055			
Db	73	TTTCCTATGTCGACAGATACACGCTTTATAGGCTGACTTCCACCTGCACCTGATAT	132			
Qy	1056	GTATCCGACTATATGATCTATTTGGTGGTGTAACGATTCGTAAATTTGGATGCTTA	1115			
Db	133	GCAAC--ATTACACATGATGCTTCTAAGTCAATACGATTTCAACAGATAGGAGGA	189			
Qy	1116	ACTGAGCCCTTAGTACATCTGTATGTTGTTCTCCTAACAGATATGTTTCTTCTGCTTT	1175			
Db	190	GTAGCTGCGAAGATACCCCAACCTTCTGTATGACCGGAGTGGCTTCTTCTCTCTG	249			

QY	1176	GTATCTTTAGAAACAAACCTCTTG6GCGTGGCGCATGTGGATGCAACCTCTCCAG	1235
Db	250	ATATCTTTTGGTGGCCCGTTTGGCACCACCTGTGTTGCCAACTTGGATGACACCGCTGCTC	309
QY	1236	CTAGCTGTGGTGTGTGGAGCTATACAGAAATATCTTTTCGAAATTCGACAAATATACGCTCTC	1295
Db	310	CGACGCTGTACGTTGGTGGCCCTTTACAGAAATATCTTTCAGCAAGAGCGCCAAAGTACAGCTTG	369
QY	1296	TTTGACTCACTAAAGAAATGGCCTATATCCCTTTGACCAAGAGCAAAAGTCAAAAGT	1355
Db	370	TTTCGATCTCTGGCAAGAAATGTGCTATATATCCATTGGATGAGACACCAAGGTTAAAGCG	429
QY	1356	AAGCGTGCATATGATGTAGTTGTCGCCGCCCTTCGGAAATTCAGGAGAGACTTAAATCAA	1415
Db	430	AAAGCTGCTATTTGATGTGTCTCTGCAACCACTTGGGAAATTCAGCGGTCTCTGATCCAG	489
QY	1416	CAAGCTTTGCTGCTATCTGTGGAAGTATTTGAGCTATGACCCCTTATCTTGACAGTAT	1475
Db	490	CAGTTTATGATCTTCACGTTTGGCTGACGTGCGCAGGCTTACATCTTACCTTGAGTCAAT	549
QY	1476	CTT 1478	
Db	550	CTT 552	
RESULT 8			
AV926113			
LOCUS	AV926113	645 bp	mRNA
DEFINITION	AV926113 K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA		
ACCESSION	AV926113		
KEYWORDS	AV926113.1 GI:18221910		
SOURCE	EST.		
ORGANISM	Hordeum vulgare subsp. vulgare.		
REFERENCE	Hordeum vulgare subsp. vulgare. Hordeum vulgare subsp. vulgare Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.		
AUTHORS	I (bases 1 to 645)		
TITLE	Sato, K., Saitoh, D. and Takeda, K.		
JOURNAL	Barley EST sequencing project in NIG and Okayama Univ Unpublished (2002)		
COMMENT	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. Location/Qualifiers		
FEATURES			
Source	1..645		
	/organism="Hordeum vulgare subsp. vulgare"		
	/cultivar="Haruna Nijo"		
	/db_xref="taxon:112509"		
	/clone="basd18c14"		
	/clone_lib="K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves"		
	/tissue_type="seedling leaves"		
	/dev_stage="second leaf stage"		
BASE COUNT	108 a 202 c 168 g 167 t		
ORIGIN			
Query Match	10.9%, Score 178, DB 10, Length 645;		
Best Local Similarity	54.8%, Pred. No. 3.9e-31;		
Matches 352; Conservative 0; Mismatches 290; Indels 0; Gaps 0;			
QY	123	AAGAAAGTTGGCCAAATGTCCTATATGTTCTTCTGATATTAACATTAATAGGCTGTTA	182
Db	2	AAGAAGATCGGCGCCCTGGTCTATGTTCTTCTGATATTCATCTTCACTACACATCTCG	61
QY	183	CGCGATACAAAGACACTTATTTGTGGAGTCTCGTGTCTGATGTCAGAGCATATCT	242

```

Db      62 AGGAGACCAAGAGCGTCTCGTCGACCGCAAGGGAGCAGCGCCAGATCATCCCT 121
Qy      243 TTCATCAAGTTTGGCTTTGTCCTGCTGCTATTTATGTTATGTAAGAAG 302
Db      122 TTCCCAAGACGTGGGTCAACCTCCCAAGGCGGCTGCTTCCTTACTCCAAAG 181
Qy      303 CTAAGTAATATTTAAGTAGAGGCGCTATTTATGACGGGAACGCGCTTTTAATT 362
Db      182 CTCGCGGACGCTCTCCACAGAGGCGCTCTTCTACCGCATCTTCCCTTCATCGCC 241
Qy      363 TTCTTGGCCCTGTCGCGACTGTAAATTTATCCGCTACGCGATGTTTATACATCCAGAA 422
Db      242 TTCCTTGGGCTCTCGGATAGCTGCTACCTATGCGGAGCGCATCCACCCAGCGCG 301
Qy      423 TTGCTGACCGTTTACAGGCCATCTACCTCCAGAGATTGCTAGACTGCTTCCATCTTA 482
Db      302 CTCGCGGACGCGCTCTCCGCTGCGCGCGCCAGCTTCTGGGGCCGCTCCATCCTTA 361
Qy      483 AGAAGCTGACATTTGCTGCATTTTATGACTTGTGACATATGGGGAAGCGTCAATGCTA 542
Db      362 CCGCTGTGAGTTTCTGCTCTTTTACGTTATGCGGAGCTCTGGGGCAGCGTGTGCTATC 421
Qy      543 TCTTAATGTTCTGGGATTTGCTATGAATTAACAAATCCAGCAAGCAAGCGTTTC 602
Db      422 TCCGTCTGTTTGGGATTTGCCAATGACATGACAGCGTTGAAAGAGCTTAAGAGTTT 481
Qy      603 TAGCGCTCTTTGGATTCGAGACTAATTTCTTACTAGCTTCTGCTGCTCAATTTGTT 662
Db      482 TAGCCACTGTGTTGGGCTTGGGCGCAATGTGCTCTGCTGCTGCGACGCGGAAA 541
Qy      663 TGGGCTTCAAGTTGAGACTTCCGTTTCTGAAGGTGTGATCCCTTGGGGAATTTCTTA 722
Db      542 TACTTCTCAAAATGAGCGAGAATTTGGGTCAGAGGGGTGAGCGAGTGGGCTATTTCATTG 601
Qy      723 GCTCTTTGAGCTATGACTATTTATCTGAGACTTGTCTT 764
Db      602 AAGGCGATGATGAGCATAGTGGTGTGCGGGTTGCTATT 643

```

```

RESULT 9
BI405734
LOCUS    115606 Mature tuber lambda ZAP Solanum tuberosum cDNA, mRNA
DEFINITION
Sequence.
ACCESSION BI405734
VERSION   BI405734.1 GI:15185148
KEYWORDS EST.
SOURCE    potato.
ORGANISM Solanum tuberosum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 664)
AUTHORS  Nielsen,K.L., Crookshanks,M., Emmersen,J. and Welinder,K.G.
TITLE    EST-sequencing of mature potato tuber (var. Kuras)
JOURNAL  Unpublished (2000)
COMMENT  Contact: Karen G. Welinder
         Institut for bioteknologi
         Aalborg Universitet
         Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
         Tel: +45 96358467
         Fax: +45 98141808
         Email: kwebio.auc.dk
         Sequenced from the 5' end.
         High quality sequence stop: 664
         POLYA-No.

```

```

FEATURES
Source    location/Qualifiers
1..664
/organism="Solanum tuberosum"
/cultivar="Field grown Kuras"
/db_xref="taxon:4113"
/clone_lib="Mature tuber lambda ZAP"

```

```

/lisse_type="Tuber"
/note="Vector: lambda ZAP"
BASE COUNT 171 a 143 c 157 g 193 t
ORIGIN
Query Match 10.3%; Score 168.4; DB 13; Length 664;
Best Local Similarity 56.4%; Pred. No. 6.8e-29;
Matches 335; Conservative 0; Mismatches 256; Indels 3; Gaps 1;

```

```

Qy      939 TATGTAATTTGATTAATTAATCAAGTACTGTGAAAGTCAGCTGAACCTGAATAT 998
Db      1 TATGGATTAATGATCAACCTGTGAAGTTATCATGAAGTCAAAAGCTCAAGCTTTC 60
Qy      999 CCTAATATGAATGACTATGATGATTCATGGGGAACCTTCTCTTGAGCTGGCGTAGTA 1058
Db      61 CCAAGCCCAATGAATATCTCATTCATGAGGTGATGATCTCAAGTCTCACTGAATATGA 120
Qy      1059 TCCGTAATATATATATATGTTGTTGTTGTAAGCAATCTGTAATATTTGATGTTAACT 1118
Db      121 ACTTTCATATATGATGTTGTTAGTCAATGSA---TTTTCGACAAAGTATGGTGGGAGCA 177
Qy      1119 GGAGCCCAATGATCTGCTGATGTTTCTCAAGAGTATGTTTCTGCTCTTCTT 1178
Db      178 GCAGCCCAATGATCAACCTGATGTTGTTCTCAACCGAGTGTGTTCTTCCCTGCTT 237
Qy      1179 ATCTTGAAGAACCAAGCTTGGGCTGCTGCTGCTATGTTGGGTACAACTCTCTCATGCTA 1238
Db      238 TTGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
Qy      1239 GCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1298
Db      298 GCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 357
Qy      1299 GACTCAACTAAGAAATGCTGATATCTCTTGAACCAAGCAAGCAAGCAAGCAAGCAAG 1358
Db      358 GACCCCTGCAAGAAATGCTGATATCTCTTGAAGCAAGCAAGCAAGCAAGCAAGCAAG 417
Qy      1359 GCTGCAATGATGATGTTGTTGCTGCAATCCATGGAAGTGTGAGAGCTTGTATACAA 1418
Db      418 GCAGCAATGATGATGTTGTTGCTGCAATCCATGGAAGTGTGAGAGCTTGTATACAA 477
Qy      1419 GCTTGTCTGTTATCTGTTGAAGTATGAGGATGAGCCCTTATCTTGCAGTATCTT 1478
Db      478 TTCATGATTTTGACTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 537
Qy      1479 CTTTTCATCATGCTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1532
Db      538 TTAGTAATGTTCTTTCATGCTGTTGGAGCAGCAAGCTTGTGATGACAGATTTC 591

```

```

RESULT 10
BE459811
LOCUS    715 bp mRNA linear EST 18-MAY-2001
DEFINITION
EST415103 tomato developing/immature green fruit Lycopersicon
esculentum cDNA clone cLEM8117, mRNA sequence.
ACCESSION BE459811
VERSION   BE459811.1 GI:9504113
KEYWORDS EST.
SOURCE    tomato.
ORGANISM Lycopersicon esculentum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
           Lycopersicon.

```

```

REFERENCE 1 (bases 1 to 715)
AUTHORS  Alcalá,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
         Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Romling,C.M.,
         Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
         S.D.
TITLE    Generation of ESTs from tomato fruit tissue, immature green
JOURNAL  Unpublished (2000)
COMMENT  Clemson University Genomics Institute

```


ACCESSION	aestlivum cDNA clone whyd4m17 5', mRNA sequence.
VERSION	BJ300874
KEYWORDS	BJ300874.1 GI:20112412
SOURCE	EST.
ORGANISM	Triticum aestivum bread wheat.
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Triticum. 1 (bases 1 to 573)
AUTHORS	Ogihara,Y., and Murai,K.
JOURNAL	Expressed genes in Triticum aestivum Unpublished (2002)
COMMENT	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.
FEATURES	Location/Qualifiers
source	1..573 <code>/organism="Triticum aestivum"</code> <code>/cultivar="Chinese Spring"</code> <code>/db_xref="taxon:4565"</code> <code>/clone="whyd4m17"</code> <code>/clone_lib="Y"</code> Ogihara unpublished cDNA library, wh_yd" <code>/feature_type="spikelet at late flowering"</code> <code>/dev_stage="Feekes", scale 6"</code> <code>/note="Vector: Lambda Uni-ZAP XR, excised phagemid;</code> Site_1: EcoRI; Site_2: XhoI; Plants were grown under hydropnic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhnov in J Dvorak Lab). Total RNA was prepared from sheath tissue, equal quantilies of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give pluescript phagemids in the TJ Close lab at the University of California, Riverside (Akhnov, Chin , Choi, Close, Fenton, Kinian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT	147 a 110 c 155 g 161 t
ORIGIN	
Query Match	10.0% Score 163.6; DB 13; Length 573;
Best Local Similarity	56.9%; Prev. No. 9e-28;
Matches 320;	Conservative 0; Mismatches 239; Indels 3; Gaps 1;
QY	841 GGAAAAAGGTCCTAAACTTAATGATGAAGAATAGCCTCTATCTTGATAGAT 900
DB	14 GGAAGACAAGAATAATGAGCAAAAGCTTAGCATGGGGAAGATTGAAGGTGCGTATCAT 73
QY	901 CTCCTTATATCTTTTATTAACCTCTCTGGTTATTTCCCTANGTATTTGCATTAACCTAA 960
DB	74 CTCATATATGTGAGGATCTTGCCACACGHCIGCTTANGTATTAAGCATCAACCTTG 133
QY	961 TCGAAGTACTTTGGAAGAGTCAGTCGAACACTGCATATCTTATATGATGATCTATAGTG 1020
DB	134 TAGAGGTACATGGAATGGAATTAAGAGGACACAGTCTCCCAAGTCTCAACGAGTATCTT 193
QY	1021 AGTTCATGGGGAACTCTCTCTCTTGAGCTGGGATGATTCGCTACTTATTCATGCTATTTG 1080
DB	194 CATTCATGGGATGATTTCTCAACGGCTACGTGGCATNG---CTACGTTTACAAATGATGTGT 250
QY	1081 TTGGTGTGAACCTATTCGTAAATTTGGATGTGTTTACTGAGACCCTAGTACTCTGTCA 1140
DB	251 TAGGAGATTTATTTCTTGAAGAAATTCGGGTGGGAGTTCGACGCTCAATCACCCCTCGAG 310
QY	1141 TGCTTCTCTAACAGATGATTCGTTTTCTTGCGCTTGTATCTTTTGAAGAACCAAGCTTCG 1200
DB	311 TTGTTCCTTGATCAGGAGTGGATTCCTTCTACCTGCTCTGTGTGGTGAGCCATGCATC 370

OY	1201	GCGTGGTCGCAATGTGGCTACAACTCCTGCACNCTAGCTGTGGTTCCGAGCTAATAC	1260
Db	371	CTTTTCTTAGCCACGTTTTGGATGACGCCCTTGCTTGCGCGCGCTTCTGTTGGGACACTAC	430
OY	1261	AGAATATTCTTTCGAATAATCCAAATAATCGCTTCCTTGAGCAACACTTAAGAATAATGCGCT	1320
Db	431	AGAACATTTTTCAGTAAAGAGTGCANAAGTCACTTTTTCGATCCTTGCAAACAATAATGGCGT	490
OY	1321	ATATCCCTCTTGACCACAGCAAAAAAGTCAAAAAGGTAAAGTCTGATTAATGATGATGTCGG	1380
Db	491	ACATTCCTTTGGATGAGACATGAAGGTCAAAAGCAAGCAAGCAATAGATGTGATGCA	550
OY	1381	CCCCGTTCCGAATAATCAGGAGG	1402
Db	551	ACCATTTGGGAAAAATCTGGAGG	572
RESULT	16		
LOCUS	AJ499428	527 bp	mRNA linear EST 09-AUG-2002D
DEFINITION	AJ499428 MTGIM Medicago truncatula cDNA clone mtgmacc120004b07,		
ACCESSION	AJ499428		
VERSION	AJ499428.1	GI:22089871	
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	Krajinsek, F., Mantley, K., Bartelsmeier, V., Meyer, F., Bartels, D., Bekel, T., Linke, B., Franken, P., Kuester, H., Perlick, A.M. and Puehler, A.		
TITLE	Detection of transcript sequences from mycorrhizal roots of the model mycorrhiza Medicago truncatula genotype Al7 - Glomus intraradices using the approach of an EST genome project based on an SSH library		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: Krajinskí F LG Molekulargenetik Herrenhaeuser Str. 2 D-30419 Hannover, Germany.		
FEATURES	Location/Qualifiers		
source	1..527		
	/organism="Medicago truncatula"		
	/db_xref="taxon:3880"		
	/clone="mtgmacc120004b07"		
	/clone_lib="MTGIM"		
	/tissue_type="mycorrhizal roots"		
	/dev_stage="3 weeks after inoculation"		
	/note="Vector: pGEM-teasy; genotype Al7; cDNA was prepared from total RNA using the SMART PCR cDNA system (Clontech) from roots harvested three weeks after inoculation with Glomus intraradices. This cDNA was used as tester in a Suppression Subtractive Hybridization (SSH). The SSH-cDNA fragments were generated using the SSH-adaptor sequences ctataagcactcatataggctgacggcggtgcggcggaagt and ctataagcactcatataggctgacggcggtgcggcggaagt (Clontech) and ligated after Suppression Subtractive Hybridization in to the pGEM-teasy vector from Promega. Plasmids containing cDNA inserts were propagated in E. coli TOP 10F' cells (Invitrogen)"		
BASE COUNT	140 a 113 c 114 g 160 t		
ORIGIN			
Query Match	10.0%; Score 163.2; DB 9; Length 527;		
Best Local Similarity	58.8%; Pred. No. 1.le-27;		
Matches 301; Conservative 0; Mismatches 208; Indels 3; Gaps 1;			
OY	927	TTGGTTATTGCCATAGATATTGGCATTAACCTAATCGAAGTACTTGGAAAAAGTACGCTG	986
Db	19	TTAGTGGTTGCTTAATGGAATCAGCATCATCTGTTGTAGGTTACATGGAAGCTAAGCTC	78

QY	884	ATGATATGAAGAAGT	AGCTTCCTCTATCTGTATAGATATCCCTTATATTTTATTA	AACT	923
Db	349	ATGACCACAATG	AGAGCTTGAAGTCTTTGGCTCTTCAAAATATATCCAGGATCTTGGC	408	
QY	924	CTCTTGGTTAT	TTCCTATAGTATTTTGCATTACCTTAATCGAAGTAC - TTGAAAAATCA	982	
Db	409	ACATTTGGTTGT	AGCATATGGCAATTAGTATCAACCTTGTGTGAATTACATGGGAAGTCATA	468	
QY	983	GCTGAACCTGC	AATATTCCTAATATGAATAGATATATGTAGTTCATGCGGAACCTTCCTT	1042	
Db	469	GCTCAAGCTCA	GTTCCTCCACAGCCCCCAATGAATATCTTCATTCATGAGGAGCATTTCTCAAC	528	
QY	1043	CTGACCTGCG	ATATCCGTACTATATCATGCTATTTTGTGGTAAACGTCATCGTA	1102	
Db	529	TGCTACTGSA	ATAGCAACTTTCACATATGATGTGTTAAGTCAATGGA --- TTTTGGACAA	585	
QY	1103	ATTTGATG	GTTAATCTGACGCCCTAGTCACTCTCTCATGTTCCTCTATACAGATATCGT	1162	
Db	586	GTATGGGGT	GGAGACGAGCCAAAGATTAACACCTTCACAGCTTTCCTTAACCGGAGTTGG	645	
QY	1163	TTTCTTCGCT	CTTGTATATCTTAGAACAACCTTCCTGGCGTGGCCGATGTTCGGTAC	1222	
Db	646	TTTTCTTTC	CCCTGCTTTTGTGTGGTGGCCCTTAGCACCCTACTCTTTCGGAAGTTGGAAT	705	
QY	1223	AACTCTCT	CATGCTAGCTGTGGTGTTCGGAGCTATPACGAATATTTTTCGAAATCCAC	1282	
Db	706	GACTCTCT	CTTAGCAGCTGTCTATGTGGGTGCATATGCAAGACATTTTCAGTAAGAAGTGC	765	
QY	1283	AAAAATA	1288		
Db	766	AAAGTA	771		

RESULT 20	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	JOURNAL	COMMENT
AW738613	AW738613	518 bp mRNA linear EST 18-MAY-2001	AW738613	AW738613	AW738613.1	GI:7647558	tomato.			
		EST340040 tomato flower buds, anthesis, Cornell University					Lycopersicon esculentum			
		Lycopersicon esculentum cDNA clone cTODN14 5', mRNA sequence.					Eudaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.			
		1 (bases 1 to 518)					van der Hoeven R.S., Bezzeredes J.L., Matern A.L., Holt I.E., Liang			
							F., Hansen T.S., Craven M.B., Bowman C.L., Rønning C.M., Nierman			
							W., Fraser C.M., Martin G.B., Giovannoni J.J. and Tanksley S.D.			
							Generation of ESTs from tomato flower tissue, anthesis			
							Unpublished (1999)			
							Contact: CUGI			
							Clemson University Genomics Institute			
							Clemson University			
							100 Jordan Hall, Clemson, SC 29634, USA			
							Email: http://www.genome.clemson.edu/orders/index.html			
							5 prime sequence.			

FEATURES	Location/Qualifiers
source	1. .518

```

/organism="Lycopersicon esculentum"
/cultivar="TM496"
/db_xref="taxon:4081"
/clone="CRODYN14"
/clone_lib="tomato flower buds, anthesis, Cornell
University"
/tissue_type="flower"
/dev_stage="anthesis"
/notes="Vector: pBluescript SK(-); Site 1: EcorI; Site 2:
XhoI; supplier: Tanksley; flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TM496). They

```

BASE COUNT 137 a were immediately frozen in liquid nitrogen and then
ORIGIN 112 c size-separated while remaining frozen." 117 g 152 t

Query Match	9.7%	Score 158.4	DB 10	Length 518
Best Local Similarity	58.2%	Pred. No. 1.5e-26		
Matches 298	Conservative 0	Mismatches 211	Indels 3	Gaps 1

Qy	927	TTGGTATTTCCGATATGGTATTTTCATTAATCTATATGAGACAGTCTGAAAGTCAGCTG	986
Db	4	TTGGTTTACCATATGSCATTTACTATACCACTGTTGAAGTTACATGGAAGTCMAAGCTC	63
Qy	987	AAACTCAATATCCATTAATATGAATGACTATATGAGTTCATGGGCACTTCTCTTTGG	1048
Db	64	AAACCTCAGTTCCCAAGCCCCCAAGCAATCTGCTCAATCATGGTGACCTTCCCACTGCT	123
Qy	1047	ACTGGCTAGTATCCGACTTATCATCATGATATTTGGTGGTGAAGCTCATTTGTAATTTT	1108
Db	124	ACTGGATATACCACTTTCACAAATGATGTTGTTAAGTCAATGGA---TTTTCACAAGTAC	180
Qy	1107	GGATGGTTAACTGAGCCCTAGTCACTCCTGTCATGTTCTCCTTAACAGTATCGTTTC	1166
Db	181	GGATGGGGACACACAGCAAGATMAACCTACAGTCTCTCTTCAACGGAGTTGGTTTC	240
Qy	1167	TTCCCTTTGTTATCTTTAGAACAACAAGCTTCGGGCTGTGGCTATGTGGGTCACT	1226
Db	241	TTCTCCCTGCTTTGTTGTTGGTGGCCCTCTACGACCTACTCTTGCAAAATTTGGAAATGACT	300
Qy	1227	CTCTCCATGCTAGCTGTGGTTGTGGAGGCTATACAGAAATATTTCTTTGGAATTCACAAA	1286
Db	301	CCCTCTCTACACACTCTATATGGGTGCAATGCAAGAACATTTTCATTAAGATGCAAG	360
Qy	1287	TAGCGCTCTTTAGTCACACTAAAGAAATGGGCTATATCCCTCTTGACCAAGAGCAAAA	1346
Db	361	TATAGTTTGTTTTACCCCTGCAAGAAATGGGCTACATTTCTTTGGATGAGACACCAAG	420
Qy	1347	GTCAAAGTAAAGCTCTATATGATGTAGTTGCCGCCGCTTGGAAAAATCAGAGAGACT	1406
Db	421	GTTAAAGGGAAGGACACCAATCGATGTTGTCTCAATCCACTGGGAAAGCTGTGAGAGACT	480
Qy	1407	TTAATCCAAACAGTTTGGCGATTATCTGG	1438
Db	481	TTGTATCAGCAATTCATGATTTTGACTTTTGG	512

RESULT 21	
BQ255402	
LOCUS	BQ255402
DEFINITION	BQ255402.1 MTNACG27MTM KYKC Medicago truncatula cDNA clone pKYKC-3F2, mRNA sequence.
ACCESSION	BQ255402
VERSION	BQ255402.1
KEYWORDS	GI:20456155
SOURCE	EST.
ORGANISM	barrel medic. Medicago truncatula
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
AUTHORS	1 (bases 1 to 686) Vandenbosch, K., Endre, G., Silverstein, K., Town, C. D., Van Aken, S., Utekhach, T., Cheung, F., Fraser, C. M., Lamplin, A. F. and Retzel, E. F.
TITLE	The Medicago truncatula 'kiloclone' set: A selection of ESTs from various libraries rearrayed for gene expression studies
JOURNAL	unpublished (2002)
COMMENT	Contact: Vandenbosch K

University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenbeeds.umn.edu

Minnesota sequence name: MTN62TRM. Alias clone name: KV2-1D1.
This sequence represents the 5' end of the cDNA insert. More
information is available at: www.medicago.org
Seq primer: 5'ggc TCT AGA ACT AGT gga TC 3'.

FEATURES

SOURCE

1. 686

```

/organism="Medicago truncatula"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pkvkc-3f2"
/clone_lib="KVKC"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
XhoI. cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the UniZap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XLOLR cells."
BASE COUNT      171 a      148 c      155 g      188 t      24 others
ORIGIN

```

Query Match 9.7%; Score 158; DB 14; Length 686;
Best Local Similarity 54.9%; Pred. No. 1.8e-26;
Matches 333; Conservative 0; Mismatches 271; Indels 3; Gaps 1;

```

OY 981 CAGCTGAAGTGCATATCTATATGATGATAGTGCATGGGAACTTCTCC 1040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 10 CAGGTCAAGTGCATGTCCTAGCCCAAGAGTACTCTTCTTATGGAGACTCTCG 69
OY 1041 TTCTGAGCTGGGAGTATCCGCTATATCCTATTTGTTGGTGTAGCTCATTCG 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 70 ACCGCAACTGG--AAATGCTACATTCATATGATGCTTAAATATATATTTGAC 126
OY 1101 AATTTGATGCTTACTAGGAGCCCTAGTCTCTGTCATGTTCTTCAACAGTATC 1160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127 AATATGATGGAGGAGTCTGCAAGATCAGACCTACAGTCTCTTGCAGCGAGTT 186
OY 1161 GTTTCTTCTGCTGTATATCTTTAGAAACCAAGCTTGGGCTGCTATGTTGGT 1220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 GGTTCCTTCTCTCATATGTTGGAGGCTCCAAATTCACCGGCTTCATCAATTTGA 246
OY 1221 ACAATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 247 ATGACTCCACTGACAGCTGATATGTTGGCCATGCAAGCAATTTTATGCAAGAT 306
OY 1281 ACAAAATAGCTCTCTTGTACTCACTAAGAAATGCTATATCCCTTTGACCAAG 1340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 GGTAGTACAGTTATGACCCCTGCAAGAAATGCTTACATTTCCCTTGATGAAGAT 366
OY 1341 CAAAAAGTCAAGAGTCTGCTATGATGATGCTGCGCCGCTTGGGAAATCAGA 1400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 367 ACTAAGTCAAGGGCAAGACCATGATGTTGTTGCAACCCATTTGGGAAATCTGA 426
OY 1401 GGAGCTTATCAACAAAGTGTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCT 1460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 GGAGCCCTTATCCAAAGTCAATGATCTTATGCTGCTGCTGCTGCTGCTGCTGCT 486
OY 1461 TATCTTGAAGTATCTTCTTTCATGCTATGCTATGCTGCTGCTGCTGCTGCTGCT 1520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 487 TACCTTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546
OY 1521 AACAAATATCTTACGAGTCTGCTTAAAGAAACAAGAGTGTGCTCAAGAAGATTGA 1580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 547 GATACACATTTACGATTTGCTGCTGAGAGAAAGACTGAGNAACAGATGGAAGAGA 606
OY 1581 GCTCTG 1587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 GCAGCTG 613

```

RESULT 22
BJ482169 637 bp mRNA linear EST 23-MAY-2002
LOCUS BJ482169
DEFINITION BJ482169 K. Sato unpublished cDNA library, strain H602 adult,
heading stage top three leaves Hordeum vulgare subsp. spontaneum
cDNA clone bnh63ml9 5', mRNA sequence.

ACCESSION BJ482169
VERSION BJ482169
KEYWORDS GI:21160630
SOURCE EST
ORGANISM Hordeum vulgare subsp. spontaneum.
Hordeum vulgare subsp. spontaneum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Trilicaceae; Hordeum.

REFERENCE 1 (bases 1 to 637)
AUTHORS Sato,K., Saitoh,D. and Takeda,K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genetics.nig.ac.jp.

FEATURES
source location/Qualifiers
1. 637

```

/organism="Hordeum vulgare subsp. spontaneum"
/strain="H602"
/db_xref="taxon:77009"
/clone="bnh63ml9"
/clone_lib="K. Sato unpublished cDNA library, strain H602
adult, heading stage top three leaves"
/tissue_type="top three leaves"
/dev_stage="adult, heading stage"
BASE COUNT      90 a      238 c      164 g      144 t      1 others
ORIGIN

```

Query Match 9.6%; Score 157.8; DB 13; Length 637;
Best Local Similarity 56.6%; Pred. No. 2e-26;
Matches 291; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

OY 112 CTCACGAGTAAAGAAAGTTCGCAATGTTCTTATGTTCTGATATATCACT 171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 CGCTACGCTCAAGAAAGATGTCGCCCTGGGTCTCATGTTCTTCATCCTTTCACT 182
OY 172 ATAGGCTGTATGCGATACAAAGACATCTTATGTTGGAGCTCCTGCTTGTGTCAG 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 ACACATCTGTAGAGACACCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 242
OY 232 AGGCATACCTTTATCAAGTTTGGCTGTGTTCCCTGTGCTATATCTTATGCTTA 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 AGATCATCCCTTCCCAAGACGTGGGTCAACCTCCCATGCGCATGCGCTTCATCTTC 302
OY 292 TTATGCAAGAGTAAATATATTTAAGTAGAGGCTATATTTATGCAAGGCAAGC 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 TCTACTCCAGAGCTGCGGAGAGCTCTCTCAAGAGGCGCTTCTTACACGCTATCTTC 362
OY 352 CTTTTTAATTTCTTGGCCCTTCCGACAGTGAATTTATCCGACAGGAGATGTTTAC 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 CTTTATCCGCTTCTGCGGCTTCTGAGATGCTGCTATGCTGCTATGCTGCTGCTGCT 422
OY 412 ATCTTACAGATTTGCTGACCGCTTACAGGCACTCTACCTCGAGAGTCTGAGACTCG 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 ACCCCACGGGCTGCGGAGACCGCTCTGCGCTGCGGCTGCGGCTGCGGCTGCGGCG 482
OY 472 TTGCAATTAAGAAATGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 483 TCGCATCTCAGCGCTGAGATTTGCTGCTTCTTACGTTATGCGGAGAGCTTGCGGCA 542
OY 532 GCGTCATGCTATCTGATGCTGCTGCGGATTTGCTATGAATTTACAAAATTCACGAAG 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 543 GCGTGTCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602

```


VERSION BM357837.1 GI:18426789
KEYWORDS EST.
SOURCE Triphysaria versicolor.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; easterids I; Lamiales; Orobanchaceae; Triphysaria.
REFERENCE 1 (bases 1 to 529)
AUTHORS Matvienko, M., Torres, M.J. and Yoder, J.I.
TITLE Transcriptional responses to the hemiparasitic plant Triphysaria versicolor to host plant signals
JOURNAL Plant Physiol. 127 (1), 272-282 (2001)
MEDLINE 21437952
COMMENT Contact: Yoder, J.I.
John I. Yoder Research Lab, Dept. of Vegetable Crops
University of California at Davis
137 Asmundson Hall, One Shields Drive, Davis, CA 95616, USA
Tel: 530 752 1741
Fax: 530 752 9659
Email: jlyoder@ucdavis.edu
High quality sequence stop: 530.
Location/Qualifiers
1..529
/organism="Triphysaria versicolor"
/db_xref="taxon:64093"
/clone_lib="Triphysaria versicolor root-tip, early
DMRQ-induced transcript cDNA library"
/tissue_type="root-tips -5mm length"
/dev_stage="3-4 weeks growth"
/lab_host="E. coli"
/note="Vector: PCR2.1 TA Cloning System, Invitrogen,
Carlsbad, CA; PCR-based suppression subtractive
hybridization cDNA library"
BASE COUNT 101 a 97 c 142 g 189 t
ORIGIN
Query Match 9.0%; Score 146.8; DB 13; Length 529;
Best Local Similarity 55.8%; Pred. No. 7.7e-24;
Matches 280; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
OY 294 TATGCAAGCTAAGTATATTTAAAGTAGAGAGCCCTATTATGAGGGAAGGCC 353
DB 28 TACACTAAGTGGCTAATGTTTGCAGAAAGATGCATTTCTACACGTATATGTGCA 87
OY 354 TTTTAAATTTCTTGGCCCTGTCGCCACTGAATTTATCCGCTACGCGATTTTACAT 413
DB 88 TTCAATGGCTTTTGGGGCATTTGGGTTTGCTTATCCCTCAGTAATTTTTCAC 147
OY 414 CCTACAGAAATTTGCTGACCGTTTACAGGCCATCTCACTCCAGAGATGCTAGACTGGT 473
DB 148 CCCACTGGGCTTGGATTAAGCTTTTGAACGTGGGTCCGAGTTCTTGGACCACTT 207
OY 474 GCCATTTAAGAACTGGACATTTGCTCATTTATGACTTGTGAACATAGGGGAGC 533
DB 208 GCTATTTTGAAGATTTGAGCTTTTGTGTTGCTATGTTAGCGTGAGCTTTGGGCGAGT 267
OY 534 GTCATGCTATGCTAATTTCTGGGGATTTGCTATGAATTTACAAAATCAGCAAGCA 593
DB 268 GTGGTGTGTTCTGTTTGGGGGTTTGCAAATTCAGTAATCACACATGACCAAAACC 327
OY 594 AAGCGTTTCTACGCTCTTTCGATACGAGCTAATATTTCTTTCATGCTTCTGGTGT 653
DB 328 AAGAAATTCATCCCTTTTGGGCTCGGAGCAATGTAGCAATCTTTCTCGGGTGG 387
OY 654 GCAATTTGTTGGGCTTCAAGTTGAGAGCTTCGTTTGAAGGTAGATCTTGGGGA 713
DB 388 ACTGTGAATTAATTTCCCAATAGAGCAAAATTTGGGCCCGGTTGATGTGTGGGCC 447
OY 714 ATTTCTTACGCTTTTATGAGCTATGACTAATGATGAGCTGTTCTATGCCCAGT 773
DB 448 ATCTGTTGAAGGAAATATGATGTTAGTGGGAGTGGCCTTATGATTTGTTTCTA 507
OY 774 TACTGTGATCAATTAAGACG 795

DB 508 TATGTGGTGAACAAATATG 529
RESULT 32
C74932 469 bp mRNA Linear EST 04-APR-2002
LOCUS C74932
DEFINITION C74932 Rice cDNA from immature leaf including apical meristem Oryza sativa (japonica cultivar-group) cDNA clone E60955_1A, mRNA sequence.
ACCESSION C74932
VERSION C74932.1 GI:2443161
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 469)
AUTHORS Sasaki, T. and Yamamoto, K.
TITLE Rice cDNA from immature leaf including apical meristem
JOURNAL Unpublished (1997)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rjp.dna.affrc.go.jp/PROJECT="RGP".
Location/Qualifiers
1..469
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone_lib="E60955_1A"
/clone_lib="Rice cDNA from immature leaf including apical meristem"
/dev_stage="Immature"
/note="Organ: Leaf; Immature leaf including apical meristem (under long day condition)"
BASE COUNT 106 a 100 c 121 g 141 t 1 others
ORIGIN
Query Match 8.9%; Score 146.2; DB 14; Length 469;
Best Local Similarity 57.8%; Pred. No. 1.1e-23;
Matches 259; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
OY 1096 TTCCGAATTTGAGATGTTAACTGAGAGCCCTAGTCTCTGCTATGTTCTCTACAG 1155
DB 5 TTAGGAAGTTTGGTGGGAGTCGAGCTACATATACCCCTGCACTGTTGCTTCTACTG 64
OY 1156 GTATGTTTTCCTGCTTGTATCTTTAGAAACCAAGCTTGGCGTGGCTATGT 1215
DB 65 GAGTTGGAATTCCTCAGATGATTTATTCGGAGACCAATGATCTCTTATGAGCACAG 124
OY 1216 TCGGTACACTCTCTCATGCTAGCTGTGTGTTGTGAGACTATACAGAAATTTCTTGA 1275
DB 125 TTGGAATGACACTTTCCTGCTGGGTCATATGTTGGGCAATGACAAATTTTCAGTA 184
OY 1276 AATCCAAATATAGCTCTCTTGTACATCACTAAGAAATAGGCCATATTCCTGTGACC 1335
DB 185 AGAGTGCAGATGATGATTTTGTGACCTTGTAAAGAAATGCAATATCTTTGGACG 244
OY 1336 AAGACCAAAAAGTCAAAAGTGAAGCTGATTTGATGATGATGATGATGATGATGAT 1395
DB 245 AGGACATGAAGGTCAAGGTGAAGCAAGCAATGATGATGATGATGATGATGATGATGAT 304
OY 1396 CAGAGGAGCTTTATCCAAAGCTTGTCTGTTATCTGTGGAAGTATTTGAGACTATGA 1455
DB 305 CTGGAGGGGCTTGTATGATGAGCTTATCATTTTGGATCTTCGCAAAATTTCTA 364
OY 1456 CCCCTATCTTGCATGATTTCTTTCATCATCTTATTTGTTGTTCTTCGCACTA 1515

```

Db 365 CACCTTACCTGGAGAGAAATTCATTTGTAATGTTCTTGCCTGGTGGCTGCTTAAGCT 424
Oy 1516 AGTTAAACAATAATCTTATGAGCAGTC 1543
Db 425 CCCTTGACTGCANAGTTCTCCACCTGGC 452

RESULT 33
LOCUS AM924705 581 bp mRNA linear EST 19-JUL-2000
DEFINITION W51.71.D05.b1_A002 Water-stressed 1 (W51) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION AM924705
VERSION AM924705.1 GI:8090530
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 581)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: water-stressed plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmprratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 528
POLYA-No.

FEATURES
source 1..581
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_id="Water-stressed 1 (W51)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: Lambda zap; Site 1: XhoI;
Site 2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 149 a 125 c 147 g 160 t
ORIGIN
Query Match 8.9%; Score 146; DB 10; Length 581;
Best Local Similarity 58.9%; Pred. No. 1.2e-23;
Matches 251; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Oy 1118 TGAAGCCCTAGCTACCTCTGATCGTTCTCCTAACAGATCGTTTCTTCGCTCTGT 1177
Db 11 TGCAGCTACGATCAACACTGCAGTGTACTCTCACAGAGTGGGTTCTTCTCAGTAT 70
Oy 1178 TATCTTTAGAAACACAGCTTTGGGCTGGTGGCTATGTTCCGTACACTCCTTCATGCT 1237
Db 71 TTTGTTGGTAGCCATTCGATCTCTTATACCAAGTTGGATGACACCTTGGCTTGC 130
Oy 1238 AGCTGTGGTGTGGAGCTATACAGAAATATCTTTCGAAATCCACAAATACGGCTCTT 1297
Db 131 GGCAGTCTATCTTGGAGCAATGCAAGACATTTTCAGTAGAGTGCAAAAGTACAGTCT 190
Oy 1298 TGACTCACTAAGAAATGGCTATATCCCTTGACCACAGCAAAAGTCAAGGTAA 1357
Db 191 TGATCTCTGCAAGAGATGCGATACATTCCTTTGGATGAGATATGAAGGTAAAGTAA 250
Oy 1358 GCGTGTATGTAGTGTAGTGGCGCCCGCTTCGAAATTCAGAGAGAGCTTAAATCAACA 1417

```

```

Db 251 AGCAGCTATATGACGTTGCTCGACACCCCTGGGAAATCTGAGCGTCTTGATCCACA 310
Oy 1418 AGCTTGCCTGCTATCTGGGAGGATGAGACTATGACCCCTTATCTTGACGTATCT 1477
Db 311 GTTCATGATCGTGTGATTTGGATGATCTTCTTCCGATTCGACACCTTCTGGTGAATCT 370
Oy 1478 TCTTTTCATTCATTCGCTATTTGGTTGGTTCTGCAACTAGTTAAACAACATTCCTAGC 1537
Db 371 TCTGTTGATTTGTTCTTGATGCGCTGGGCTGTAAAGTCCCTGACATCAGATTTCAAC 430
Oy 1538 GCAGTC 1543
Db 431 CCTGGC 436

RESULT 34
LOCUS BG322472 522 bp mRNA linear EST 27-FEB-2001
DEFINITION EML_28.D05.b1_A002 Embryo 1 (EML) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BG322472
VERSION BG322472.1 GI:13152150
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 522)
AUTHORS Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmprratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 505
POLYA-No.

FEATURES
source 1..522
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_id="Embryo 1 (EML)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 130 a 110 c 133 g 149 t
ORIGIN
Query Match 8.8%; Score 144.6; DB 12; Length 522;
Best Local Similarity 59.3%; Pred. No. 2.5e-23;
Matches 246; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

Oy 1129 TCACCTCCGTCATGCTTCTCCTAACAGGATCGTTTCTTGGCTCTGTATCTTGGAA 1188
Db 15 TCACCTCGCAGCTGTACTTCTCACAGAGATGGGCTTCTCTCACTATTTTGTGGTG 74
Oy 1189 ACCAAGCTTCGGGCTGGCTGCTATGTTGCGTACACTCTCTCATCTAGCTGTGGTTG 1248
Db 75 AGCATTTGACCTCCTTATGACCAAGTTGGGATGACACCTTTCCTTGGCGCATCTATG 134
Oy 1249 TCGAGCTATACAGAAATATCTTTCGAAATCCACAAATACGCTCTCTTGAGCTCAACTA 1308

```

QY	571	AAATTCACAAAATACACGAAACCAAGGTTTCTACGCTCTTTTCGATATGGGAGCTAATA	630
Db	1	AGATCAGCAGCTGTGGATGAGGCTTAAGGAAATTTCTATCCTTGTGTGGACCTTGAGACGAATG	60
QY	631	TTTCTTTACTAGCTCTTCGCTGTCGCAATTTGTTGGGCTTCGAAATGTAGAGCTTCCGTT	690
Db	61	TTGCTCTTATTTTCTCTGGTGCACAGTGAAGTACTTTTCTAGCTGTAGAGGCTCTTTAG	120
QY	691	CTGAGAGGTGTGATCTGTGGGGAATTTCTTACGCTTTTGTAGTGCATGACTATGTAT	750
Db	121	GTCCCTGAGTTGATGGTTGGGCCATTCCTCCGGAAGAAATGATAGCATTTGTGTGATGA	180

```

/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="c5FB34M17"
/clone_1b="potato leaves and petioles"
/tissue_type="leaflets and petioles"
/dev_stage="8 weeks old plants"

```


Db 400 GATCTTAACCTTTGGTCTACTAGTAACCTCCGCTAGGAGGGCTCTTCTGCT 459

QY 1484 CATCATTCGTATTGGTGGTTCCTGCACCTAGTTAAACAACATCTTCTAGCCAGTC 1543

Db 460 GATTCTTCTTGGCTGATATCAGCAACCAAGTCGCTGTATACAGATTACTGCTTGG 519

QY 1544 TGCCTTAAAGAACAGAGTGCCTCAGAGAAATTCAGTCTCTGCT 1589

Db 520 TCGAAGAGAGAACTTGAGATAGAGATGAGAGACGACGCTGCT 565

RESULT 38

BF423978 606 bp mRNA linear EST 06-DEC-2001

LOCUS BF423978

DEFINITION ssc2f04.y1 Gm-cl047 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl047_1784 5', similar to SW:TIOL1.SOLIU 024361 PLASTIDIC

ACCESSION BF423978

VERSION BF423978.1 GI:11411967

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 606)

AUTHORS Shoemaker, R., Kelm, P., Vodkin, L., Eipelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, E., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Mesterion, R. and Wilson, R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R/public soybean EST project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccluresgen.com
High quality sequence stop: 443.
Location/Qualifiers

FEATURES

source

1. 606

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl047-1784"

/clone_lib="Gm-cl047"

/tissue_type="Immature leaves (unfurled trifoliate) of greenhouse grown plants"

/dev_stage="2 week old"

/lab_host="DH10B"

/note="vector: pSPORT1; Site_1: NotI; Site_2: SalI; This cDNA library was constructed from mRNA isolated from immature leaves (unfurled trifoliate) of greenhouse grown plants that were 2 week old. The library was prepared using the Life Technologies psupertscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linker adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed in the laboratory of Dr. Lilia Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign (email l-vodkin@uiuc.edu)"

BASE COUNT 156 a 131 c 144 g 175 t

ORIGIN

Query Match 8.7%; Score 143; DB 12; Length 606;
Best Local Similarity 55.8%; Pred. No. 5.9e-23;
Matches 294; Conservative 0; Mismatches 230; Indels 3; Gaps 1;

QY 975 AAAAGTCACCTGACCTGATATCTATATGATAGATAGTACATGAGTTCATGGGAGC 1034

Db 1 AATATCAAGCTCAAGAGCTAGTTTCTTACGCCCAAGATGACTCATCTTTATGGGAGC 60

QY 1035 TTCTCCTTCTGAGCGGCTAGTATCCGCTATATCATCTATATGTTGGGTGTAACGTC 1094

Db 61 TTCTCAACTGCAACGGAATGCTACATTCATTAATGATGTC--TTCTAGTCAATTATTA 117

QY 1095 ATTTCGTAATTTGATGTGTTAACTGAGCCCTAGTACCTCTGTCTATGGTTCCTTACA 1154

Db 118 TTTGACAAATATGATGAGGAGGTGGTGCAGATCACCCCAACGCTCTCTGATATGACA 177

QY 1155 GGTATCGTTTCTGCTGCTCTGTTATCTTACGAACCAAGCTTGGCGGTGGCTATG 1214

Db 178 GGAGTTGGTTTCTTTCTTCTTATATATTGCGGCCCAATGCTCTTATTGCAAG 237

QY 1215 TTGCGTACAACTCTCTCATGCTAGCTGTTGTCGAGCTATACAGATATTTCTTGC 1274

Db 238 TTTGGAATGACTCCACTGCTAGCTGCTGTATATGTTGGGCCATGCAAGAAATATTCAGC 297

QY 1275 AATTCACAAATATACGCTCTTTGACTGACATAAAGAAATGCGCTATTCCTCTTGAC 1334

Db 298 AAGAGTGCTAATGATACGTTTATTCGATCCGTCGCAAGAAATGCGCTATACCTTAGAT 357

QY 1335 CAAGAGCAAAAGTCAAGAGTAAAGCTGTATGATGTTGGCGCGCTTGGCAAAA 1394

Db 358 GAGGATACCAAGCTCAGAGAAAGGACCATGATGATGTGTGTCAGCAACCATGAGAGAA 417

QY 1395 TCAGAGAGACTTATATCCACAAAGTTCCTGTTATCTGTGAGATTTGAGCTATG 1454

Db 418 TCTGGAGAGACTGCTGATTCAGAGCTCATGATCTTATGCTTGTGCTACATGACTACA 477

QY 1455 ACCCGTATCTTGCACTGATTTCTTTTTCATCATTTGCTATTTGGTT 1501

Db 478 GCCCATACCTAGGAGGGGTCTCTGTGATGTTCTTGGCGGT 524

RESULT 39

BF456211 663 bp mRNA linear EST 19-MAR-2001

LOCUS BF456211

DEFINITION NF075B07P.L1059 Phosphate starved leaf Medicago truncatula cDNA clone NF075B07P.L 5', mRNA sequence.

ACCESSION BF456211

VERSION BF456211.1 GI:13379536

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE 1 (bases 1 to 663)

AUTHORS Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.

TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula phosphate-starved leaf library

JOURNAL Unpublished (2000)

COMMENT Contact: Harrison MJ
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mjharrison@noble.org
Insert Length: 663 Std Error: 0.00
Plate: 075 row: B column: 07
Seq primer: TCACACAGAAACACGCTATGAC.

FEATURES	Source	Location/Qualifiers
		1. .653
		/organism="Medicago truncatula"
		/db_xref="taxon:3880"
		/clone="NF075807PL"
		/clone_lib="phosphate starved leaf"
		/tissue_type="leaf"
		/dev_stage="trifoliolate"
		/note="Vector: Lambda Zap: At the trifoliolate stage, M. truncatula plants were transplanted to phosphate-free sand and grown for a further 30 days. During this 30 day period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 200m potassium phosphate. RNA was prepared from above ground tissues."
BASE COUNT	181 a	140 c 156 g 186 t
ORIGIN		
Query Match	8.7%	Score 142.8; DB 12; Length 663;
Best Local Similarity	54.9%	Pred. No. 6.6e-23;
Matches	282; Conservative	0; Mismatches 232; Indels 0; Gaps 0;
Qy	1074	CAATTTGGTGGGATACGCTCATTCGTAAATTTGGATGGTGTACTGAGACCCCTAGCACT 1133
Db	4	CGAGTTTCTAAGTCAATATATATATTGACAAATATGGATGGGAGTTGGCTGCCAAGATCACA 63
Qy	1134	CTGTCATGTTGGTTCTCTCTTAACAGGATATCGTTTTCCTGCGCTTGTATCTTTATGAAACAA 1193
Db	64	CCTACAGTCCTGCTTCTGTGACCGGAGTTGGTTTCTTCTCTCATATATGTTGGAGTCCA 123
Qy	1194	GCTTTGGGCTGGTGGCTATGTTGGGTACACTCTCTCATGCTATGCTGTGGTTGGGA 1253
Db	124	ATTGCAACCCGGCTTGGATCATCAATTTGGAAATGACCTCCACCTGTAGCAGCTGTATATCTTGCT 183
Qy	1254	GCTATACAGAAATATTTCTTTCGAAATTCACAAATTCAGCGCTCTGTAGCTCACTAAAGAA 1313
Db	184	GCCATGCGACAACTTTTATAGCAGAGTGTACTAGTATATTTGACCCCTGCACAAAGAA 243
Qy	1314	ATGGCCTATATCCCTCTTGACCAAGAGCAAAAAGTCAAAAGGTAAAGCTGTATGATGTA 1373
Db	244	ATGGCTTACATTTCCCTTGGATGATAGATACTAGTGCACAGGCGGAAAGCAGCATTTGATGTT 303
Qy	1374	GTTGGCCGCCCGCTTGGGAAATCAGAGAGAGCTTTAATCCAAACAGTTTGGTGGTTATC 1433
Db	304	GTTTCAACACCCATTTGGGAAATCTGAGAGAGCCCTTATCAACAGTTCATCTTAACT 363
Qy	1434	TGTGAAGATATGAGAGCTATGACCCCTTATCTTGGAGATGATCTTTCATTCATTCCT 1493
Db	364	TTTGGTTACTATGCAACTCACTCATTCATCTTGGAGGTTGGTCTTCTTGTGATTTGTTCTT 423
Qy	1494	ATTGTTGGTTGTTTTCGCAACTAAGTTAAACAACATATCTTAAAGCCAGCTGCTCTTAAA 1553
Db	424	GCGCTGTTAAGGAGCAGCCAGCAAGTCTTTGGATACACAGTTTACTGCAATTCGCTGAGAGGAA 483
Qy	1554	GACCAAGAAGTGGCTCAAGAAGATTCAGCTCGT 1587
Db	484	GAACTCGAGAAACAGATGAAAGACAGCAGCAGCTG 517
RESULT 40		
BE726529		
LOCUS	BE726529	575 bp mRNA linear EST 14-SEP-2000
DEFINITION	894092JF04.y2 C. reinhardtii CC-1690, normalized, lambda Zap II	
ACCESSION	BE726529	Chlamydomonas reinhardtii cDNA, mRNA sequence.
VERSION	BE726529.1	GI:10127825
KEYWORDS	EST.	
SOURCE	Chlamydomonas reinhardtii.	
ORGANISM	Chlamydomonas reinhardtii	
REFERENCE	1 (bases 1 to 575)	
AUTHORS	Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., Mcdermott,J.P., Silflow,C., Stern,D. and Surzycki,R.	

TITLE	Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants: Project phase 2				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Charles Hauser DCMB Box 91000 Duke University Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chausser@duke.edu				
FEATURES	Location/Qualifiers				
source	1..575 /organism="Chlamydomonas reinhardtii" /strain="CC-1690 Wild type mt+ 21gr" /db.xref="taxon:3055" /clone_lib="C. reinhardtii CC-1690, normalized, lambda zap II" /note="Vector: pluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in YAP (acetate-containing) medium in the light, YAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites pluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsistit (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."				
BASE COUNT	103 a 201 c 158 g 113 t				
ORIGIN					
Query Match	8.6% Score 141; DB 12; Length 575;				
Best Local Similarity	54.1% Pred. No. 1.7e-22;				
Matches 310; Conservative	0; Mismatches 260; Indels 3; Gaps 1;				
QY	855	AAACCTAAATAATATGAAGAATAGCTCCCTATCTGTAGATGATTCCTATATTCWT	914		
DB	5	AAGACCMAGATGAGCTGGGGCAGTCTTCCTGCGCCAGTCAGCTACATCCGC	64		
QY	915	TTATTAACCTCTTGATTATTTGCCATATGATGATTTGCATTAATCAAGAACTTGG	974		
DB	65	GACATGGCCACCTGATGGTGCCTACGGTATCTGCATCAACCTGATGAGAGTACCTGG	124		
QY	975	AAAAGCTACGCTAAACTGCATATATCTTATATGATGACTATATGAGATCTTACGGGAAC	1034		
DB	125	AAGGGCAAGCTCAAGGCCACAGTTCCCAACCCCAACGACTCTGCTCTGGGGCAG	184		
QY	1035	TTCTCTTCTTGAGCTGGGGGTAGTATTCGCTATATCAATGATCTTTTGGTGGTAAACGC	1094		
DB	185	TTTCTCACCCGCCACCGTACCTGATCCTTACCAATGATGATCTTCTCCCGCTGGATCTTC	244		
QY	1095	ATTCTAAATTTGGATGTTAACTGGAGCCCTAGTCACTCTGTCATGATGTTCTCTTACA	1154		
DB	245	AAGC---AGTTGGCTGGGGGTGTGGCTGCCGATACACCCCATCATGCTGTGATCAAC	301		
QY	1155	GGTATTCGTTTTCTTGCGCTCTTGTATTTCTTTAGAAACAAGCTTCTGGGCTGGCTATG	1214		
DB	302	GCTTCGCTGTTCTTCAACCTTGCTGGCGGGGAGACCCGTCMAAGCCCGCTTGCGCCGC	361		
QY	1215	TTCCGTTACACCTCCCTCATGCTACTCTGGTTGGTGGAGGCTATACAGATATCTTTGCG	1274		
DB	362	TGGGCGATGACCCCTCTGACGCCCGCGGCTGGTGGGGGCGGCCAGAGACTCTTCTCC	421		
QY	1275	AAATCACAATAATACGCTCTCTTTACTCAACTAAGAANAAGCCCTATATCCTCTTGA	1334		
DB	422	AAGTGTCTCAAGTACTCTCCTGTTCGACCCCTCGCAAGAGATGGCTTACATCCCCCTGGAG	481		
QY	1335	CAAGAGCAAAAAGTCAAGAGTAAAGGCTGCTATTGATGTAGTTGGCCGCCGCTTGGAAA	1394		
DB	482	GACGAGGTCAAGACCAAGGGCCAGGCCCATTTAGACGTGATTCGAACCCGCTGGGGCAAG	541		

us-09-869-433-1.rst

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 19:36:33 : Search time 2952 Seconds
(Without alignments)
16138.649 Million cell updates/sec

Title: US-09-869-433-1

Perfect score: 1637

Sequence: 1 gaataaanaactacagaa.....gttgatccctactgctttt 1637

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_or:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vtl:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB ID	Description
		Match	Length		
C 1	1535	93.8	13389	1 AE002202	AE002202 Chlamydia
2	1484	90.7	11421	1 AE001619	AE001619 Chlamydia
3	1484	90.7	300650	1 AP002546	AP002546 Chlamydia
C 4	45	2.7	23773	1 AE002301	AE002301 Chlamydia
5	39	2.4	1587	1 TAJ10586	TAJ10586 Chlamydia
C 6	39	2.4	14168	1 AE001281	AE001281 Chlamydia
7	21	1.3	5032	6 AX281684	AX281684 Sequence
8	21	1.3	8220	6 AR064709	AR064709 Sequence
9	21	1.3	8220	6 AR089156	AR089156 Sequence
C 10	21	1.3	17462	9 AL592487	AL592487 Human DNA
11	21	1.3	19124	3 PFPAVARIA	PFPAVARIA
12	21	1.3	19124	6 AR089157	AR089157 Sequence
C 13	21	1.3	90106	2 AC007499	AC007499 Homo sapi
14	21	1.3	98286	2 AC092392	AC092392 Homo sapi
C 15	21	1.3	126996	2 AC113803	AC113803 Homo sapi
16	21	1.3	134482	9 AC007335	AC007335 Homo sapi
C 17	21	1.3	143279	9 AC104838	AC104838 Homo sapi
18	21	1.3	150495	2 AC046148	AC046148 Mus muscu
C 19	21	1.3	150720	9 AL513016	AL513016 Human DNA
20	21	1.3	138599	2 AC101941	AC101941 Mus muscu
C 21	21	1.3	165679	2 AC128538	AC128538 Homo sapi
22	21	1.3	171747	9 AP001554	AP001554 Homo sapi
C 23	21	1.3	173595	9 AC020766	AC020766 Homo sapi
24	21	1.3	206463	2 AP000834	AP000834 Homo sapi
25	20	1.2	1770	6 AX098422	AX098422 Sequence
26	20	1.2	2110	8 AF428316	AF428316 Arabidops
27	20	1.2	2151	8 AY045903	AY045903 Arabidops
28	20	1.2	2181	8 ATANTMR	249227 A. thaliana
C 29	20	1.2	2762	9 AF222985	AF222985 Homo sapi
30	20	1.2	7291	9 AF222983	AF222983 Homo sapi
C 31	20	1.2	15002	9 AF222981	AF222981 Homo sapi
32	20	1.2	33376	9 AF222987	AF222987 Homo sapi
33	20	1.2	33780	6 AX151213	AX151213 Sequence
34	20	1.2	48546	9 AC107059	AC107059 Homo sapi
C 35	20	1.2	57070	9 AL354721	AL354721 Human DNA
36	20	1.2	58810	2 AC113541	AC113541 Mus muscu
37	20	1.2	65259	2 AC113102	AC113102 Mus muscu
C 38	20	1.2	74010	2 AC114147	AC114147 Rattus no
39	20	1.2	74274	2 AC087717	AC087717 Homo sapi
40	20	1.2	77422	3 AC025723	AC025723 Caenorhab
C 41	20	1.2	86897	9 AC093843	AC093843 Homo sapi
42	20	1.2	95537	2 AC021512	AC021512 Homo sapi
C 43	20	1.2	103984	2 AC129468	AC129468 Rattus no
44	20	1.2	104289	2 HS104D14	HS104D14
C 45	20	1.2	111963	2 AC099180	AC099180 Rattus no
46	20	1.2	120387	10 AL604046	AL604046 Mouse DNA
C 47	20	1.2	121637	8 AC122544	AC122544 Genomic s
48	20	1.2	122897	2 AP001953	AP001953 Homo sapi
C 49	20	1.2	127491	8 AC124962	AC124962 Medicago
50	20	1.2	132990	8 AC018848	AC018848 Arabidops
C 51	20	1.2	133968	9 AL136171	AL136171 Human DNA
52	20	1.2	136284	9 HS53C18	HS53C18
C 53	20	1.2	139505	9 HS591C20	HS591C20 Human DNA
54	20	1.2	145052	10 AL603889	AL603889 Mouse DNA
C 55	20	1.2	145616	2 HS108K11	HS108K11
56	20	1.2	147517	2 AC102030	AC102030 Mus muscu
C 57	20	1.2	150681	2 AC016535	AC016535 Homo sapi
58	20	1.2	150681	2 AC006011	AC006011 Homo sapi
59	20	1.2	150902	2 AC098652	AC098652 Homo sapi
60	20	1.2	152118	2 AC013263	AC013263 Homo sapi
C 61	20	1.2	156255	9 AP002369	AP002369 Homo sapi
C 62	20	1.2	157527	9 AC117746	AC117746 Homo sapi
63	20	1.2	158524	2 AL162372	AL162372 Homo sapi
C 64	20	1.2	159991	9 AC093628	AC093628 Homo sapi
65	20	1.2	165739	9 AL353728	AL353728 Human DNA

Pred. No. is the number of results predicted by chance to have a

C 66	20	1.2	165951	9	AC010730	Homo sapi
C 67	20	1.2	166308	9	CNS01DX2	AL139920 Human chr
C 68	20	1.2	171782	2	AC095121	Rattus no
C 69	20	1.2	171847	9	AL590229	AL590229 Human DNA
C 70	20	1.2	179044	2	AC073845	AC073845 Homo sapi
C 71	20	1.2	179391	2	AC115195	AC115195 Rattus no
C 72	20	1.2	179484	2	AC098070	AC098070 Rattus no
C 73	20	1.2	180210	2	AC109156	AC109156 Mus muscu
C 74	20	1.2	182653	2	AC130220	AC130220 Mus muscu
C 75	20	1.2	182915	2	AC019213	AC019213 Homo sapi
C 76	20	1.2	186493	2	AC125550	AC125550 Rattus no
C 77	20	1.2	187461	9	AL590453	AL590453 Human DNA
C 78	20	1.2	187490	9	AC023392	AC023392 Homo sapi
C 79	20	1.2	187638	2	AC118273	AC118273 Homo sapi
C 80	20	1.2	196514	9	CNS05BT6	AL163952 Human chr
C 81	20	1.2	197592	2	AC026232	AC026232 Mus muscu
C 82	20	1.2	198280	9	AC092506	AC092506 Homo sapi
C 83	20	1.2	198784	10	AC091521	AC091521 Mus muscu
C 84	20	1.2	199514	10	AC068902	AC068902 Mus muscu
C 85	20	1.2	205284	2	AL732520	AL732520 Mus muscu
C 86	20	1.2	209346	9	AC026227	AC026227 Homo sapi
C 87	20	1.2	210516	2	AC091274	AC091274 Mus muscu
C 88	20	1.2	219054	2	AC068663	AC068663 Mus muscu
C 89	20	1.2	246865	2	AC074003	AC074003 Homo sapi
C 90	20	1.2	298216	2	AC006875	AC006875 Homo sapi
C 91	20	1.2	346540	2	AC107420	AC107420 Homo sapi
C 92	19	1.2	155	11	664013	G64013 2156SP6 Hum
C 93	19	1.2	626	3	AF340093	AF340093 Biorhiza
C 94	19	1.2	626	3	AF340096	AF340096 Biorhiza
C 95	19	1.2	840	1	AHEEXA	X77265 A. hydrophil
C 96	19	1.2	985	11	CNS06KOP	AL100315 T7 end of
C 97	19	1.2	1001	6	AX081559	AX081559 Sequence
C 98	19	1.2	1001	6	AX374740	AX374740 Sequence
C 99	19	1.2	1185	5	VDMT0G	L33397 Valltur gryp
C 100	19	1.2	1186	5	GYMT0G	L33379 Gylmogyps c

ALIGNMENTS

source	
gene	1..13389
CDS	/organism="Chlamydomophila pneumoniae AR39"
	/strain="AR39"
	/db_xref="taxon:115711"
	/note="synonym: Chlamydia pneumoniae AR39"
gene	163..339
CDS	/gene="CP0400"
	163..339
	/gene="CP0400"
	/note="hypothetical protein; identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=1
	/product="hypothetical protein"
	/protein_id="AAF38245.1"
	/db_xref="GI:7189325"
	/translation="MLSVSMKLIIPQDSIERETSDRKDKITFTYICSSKYLACHFFSS HLDKHNKIHESIGV"
gene	597..1448
CDS	/gene="CP0401"
	597..1448
	/gene="CP0401"
	/note="hypothetical protein; identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=1
	/product="hypothetical protein"
	/protein_id="AAF38246.1"
	/db_xref="GI:7189326"
	/translation="MVTNPQVYRNQNVNSQATQFSCQAPLSLIIVSVAVAATATYA LVGSSLSLIEIGTALVYSLILPASAFMYTKKROEKKLIPKKTMLIQEHPYSI VDFTRDQEVPIYEHHLIISLUNKTNVDFKAPVLDQEKLDQFGEKFRDHPSPFPE EILLQDPLMHGRLVPMVSDVPTPGYGGWGPGLGVENAPSLERSLTLTKKI SFGEEALLDGIKKNTWSSSELVQIRQLFTFRYAADEEVEDAEINADYEQDFSLHL IFSHKLS"

AE002202/C	RESULT 1
LOCUS	AE002202/2
DEFINITION	AE002202 13389 bp DNA linear BCT 30-MAY-2000
ACCESSION	genome: Chlamydia pneumoniae AR39, section 33 of 94 of the complete
VERSION	AE002202 AE002161
KEYWORDS	AE002202.2 GI:8163425
SOURCE	
ORGANISM	Chlamydia pneumoniae AR39.
REFERENCE	Chlamydia pneumoniae AR39
AUTHORS	Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
	1 (bases 1 to 13389)
	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
	White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
	Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
	Bowman,C., Dodson,R., Gwin,M., Nelson,W., Debby,R., Kolonay,J.,
	McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE	Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
	pneumoniae AR39
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE	20150255
PUBMED	10684935
REFERENCE	
AUTHORS	2 (bases 1 to 13389)
	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
	White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
	Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
	Bowman,C., Dodson,R., Gwin,M., Nelson,W., Debby,R., Kolonay,J.,
	McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE	Direct Submission
JOURNAL	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
COMMENT	Medical Center Dr, Rockville, MD 20850, USA
FEATURES	On Jun 1, 2000 this sequence version replaced gi:7189324.
	Location/Qualifiers

gene 1455..1808
/gene="CP04.02"
CDS 1455..1808
/gene="CP04.02"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38247.1"
/db_xref="GI:7189327"
/translation="MQLFYNNESGMDLCPDSQGBGFQLRLVGLHSSMAIYEAKE
EFTYLPKVSLLTWELIMQLSKPTKRGVAKDCNVFEKHFQRFROYLGLSLDINRF
ENFTFNPKRYHLDRD"
1842..1958
/gene="CP04.03"
1842..1958
/gene="CP04.03"
/note="CP04.03"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38248.1"
/db_xref="GI:7189328"
/translation="MKSLSVRSAFFSGIVKSLPNLKGKRELFEMGFPTGNS"
2019..3326
/gene="CP04.04"
2019..3326
/gene="CP04.04"
/note="CP04.04"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38249.1"
/db_xref="GI:7189329"
/translation="MAMSMRTVPHALFNHCCHSTPELPSRTIVRTAIALSPCLGAL

AAAGCAGPVSIVGSLVLAIVLLALIFGEKKLPPTPIIDPRTNVIDEA
YGLSISAFREOVYLAIEPQFSTALCNISPEEKIKOLPELSKVESFGISLAGD
LEKNMPLFEEDLSOTCPYLWLOKIFISAGDPOVDLEPRECGYVWGLSTAK
ATICKETHHTLOOTLKEDVLLKNAKAEQMDDEVAIYERITTTAGTILKTEA
GGLIKRETISKLELLSLSTGYSPDLOLITQLEPRADWMLCEVDNSTAINLDCALVGA
LSSOULDESSIDEDVNLGLYVIDLKEAVOAFSADPKRELGLFLLRHLDSVSKRL
BSYRQGHRIALIEHGNARARYDVNFVTGARIRHKTSIFPKD"
complement(3404. .4795)
/gene="CP0405"
complement(3404. .4795)
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF38250.1"
/db_xref="GI:7189330"
/translation="MLNLRVMTQIDPRKMSNITSPTVIONNSCNYFEELKNSYTH
IVISAILLGLALIAFLVAAVSYLSGALLGLLIGVILGIIKITPMISKEO
VPEQVLRIRAHYPKPVSDFVEAKPKLKDLSIFDILNLSFVSGSSSTNYSSEL
OOKIDFEGIRAKLNEVPTASLKLESASRPPLPSLAKILOKYPFPMLGEFISAG
SKVBLHRVKKIGSLDESDSYIKPEMLPTWLPIDFPTNSSILNLHTLVLARVL
TRDYFOHLKYALANGEMNLSDNTMKQOLFARYHAAYQSYKHLSPLOEDXYNL
LDCIFKRYSWKMSLITKTPADLMENLCLDHTGRDOMEFASLIGLTYOGLIH
KESEAFSLSLIDOFKTIROSTINAMLEMLATNHSFRLPPTVHPLKRSVF
SOPEDDESSLIG"
complement(4776. .5021)
/gene="CP0406"
complement(4776. .5021)
/gene="CP0406"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF38251.1"
/db_xref="GI:7189331"
/translation="MRFRNKSLIFIKIRIYSOSGKEOKARPPFKSITSLVILL
LEAIFNEFSSITIONFNKRNKKNKINIRLFFVFTT"
complement(5081. .6361)
/gene="CP0407"
complement(5081. .6361)
/gene="CP0407"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF73665.1"
/db_xref="GI:8163426"
/translation="MRHSGCTMKCSPLTLVHIFLKNDCBCHRSCLKIRIARILIG
LVNLVLSLSEVFLAAPTISVAIGSTLAAAVILLITLVALLKSKVLPINELOKI
LYNRYPKVEFVETHTSLTVNELKIFINCKMSGDPLPLKHKAEAFIDILKSIDL
LFPEFEELILONCPYLWLSHFIDKTESVAGISGLNKQVYGLGLAPLHFGYTTIFH
SYTRPLTLISOSYKFLYSKASKNOMSPVKTCCEITFELEPNMIFRQDVGISO
FLLFESGHTWBOAKVULEKVESPMHPSALVOKICVNTTHHONLKRMOEVRNT
SSNPEPYNPTMOBKVULEKVESPMHPSALVOKICVNTTHHONLKRMOEVRNT
SSOWTSSLPOYAFHAQYKLEKIESLPIRSSL"
complement(6479. .8026)
/gene="CP0408"
complement(6479. .8026)
/note="similar to GB:M28816 SP:P19568 PID:152470
GB:AJ23269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ADP, Arp carrier protein"
/protein_id="AAF38252.1"
/db_xref="GI:7189332"
/translation="MTKTEEPKGLNSLWPIHTHELKVLPMFLMEFCITFNTYVL
RDRDVLIVAGPAGAAIPIKIFMLVWPCAIIFMLIYAKLSNLSKALFEYAAGTPE
LIFALFPYIVYLRDVLHPTFEADRLQALIPGLLGLVAILRMWTFAAFYVLAELMG

SVMLSMFGEFANETTRKTHEAKREFALFGICANISLASGRAIYVASKLRASVSGVD
PMGISRLIMMTITVSGIWMASWIMKKNVLTDRFVNPEMCKRGKAPKMKMD
SFLYIARSPYIILLALIVATIGACINLEIVYKSKLQIKQPMNDYSEFNGFSWTG
VSVYILMLFVGANVLRKGLVGLVTEVWMLTGTIVEFALYIRFNOASGVAMFGTT
PLMLAVVGAIONISKSTKATVLDSTEMAYTIPIDODOKYKGAIDIVARFGSG
GALTOOGLIGSIGANTPYLAVILFIIAIWIVSATRLNKLFLAQALKEOVAOE
DSAPASS"
complement(8227. .8742)
/gene="CP0409"
complement(8227. .8742)
/gene="CP0409"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38253.1"
/db_xref="GI:7189333"
/translation="MGFMAVROSHIKKEIEKLGKAIKRVCGKNKENDLCRYLPGPSGG
YMHHTTKKMSAPAEOLLKMLKFTLESPTPTINPRPAPRSGKRRDPTINTKT
IERVLELARQYGDLDLARSFPRKRLTSLKRELLRSINRNGIVSELNAYVEAVKAVS
SPNLEVTSPEV"
complement(8910. .9113)
gene
Query Match 93.8%; Score 1535; DB 1; Length 1389;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GAAATAAAAACTATCAGATGAAAAATGAAATATTCAGAGGTAATATGACAAAAA 60
DB 8076 GAAATAAAAACTATCAGATGAAAAATGAAATATTCAGAGGTAATATGACAAAAA 8017
QY 61 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTTCTTGGCCGATACATCTACGAGC 120
DB 8016 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTTCTTGGCCGATACATCTACGAGC 7957
QY 121 TAAAGAAAGTTCTGCCAATGTCCTAATGTCCTGATGATTAATTAATTAATGAGTGT 180
DB 7956 TAAAGAAAGTTCTGCCAATGTCCTAATGTCCTGATGATTAATTAATTAATGAGTGT 7897
QY 181 TACGGGATACAAAGACACTTATGTTGGAGACTCTGTTCTGTGTCAGAGGCAATATC 240
DB 7896 TACGGGATACAAAGACACTTATGTTGGAGACTCTGTTCTGTGTCAGAGGCAATATC 7837
QY 241 CTTTCATCAAGTTTGGCTTGTGTCCTGTGCTATTAATCTTAATCTTAATATGCAA 300
DB 7836 CTTTCATCAAGTTTGGCTTGTGTCCTGTGCTATTAATCTTAATCTTAATATGCAA 7777
QY 301 ACCTAAGTAATTTTAAGTAAGCAAGCGCTTATTTATGCACTGGGAACGCCCTTTTAA 360
DB 7776 ACCTAAGTAATTTTAAGTAAGCAAGCGCTTATTTATGCACTGGGAACGCCCTTTTAA 7717
QY 361 TTTTCTTGGCCCTGTTCGCCGACTGTAATTTATCCGCTACGAGTGTTTACATCTACAG 420
DB 7716 TTTTCTTGGCCCTGTTCGCCGACTGTAATTTATCCGCTACGAGTGTTTACATCTACAG 7657
QY 421 AATTTGCTGACGCTTTACAGGCCATCTACCTCCAGAGATGCTAGAGACTGTTGCCATCT 480
DB 7656 AATTTGCTGACGCTTTACAGGCCATCTCCACAGAGATGCTAGAGACTGTTGCCATCT 7597
QY 481 TAGAAGACGAGCATTTGCTGATTTTATGCTAGTGAAGTATGAGGGAACGCTATGCG 540
DB 7596 TAGAAGACGAGCATTTGCTGATTTTATGCTAGTGAAGTATGAGGGAACGCTATGCG 7537
QY 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAAATCCACGAGCAAGCGCTT 600
DB 7536 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAAATCCACGAGCAAGCGCTT 7477
QY 601 TCTACGCTCTTTGCGTATGAGGCTAATATTTCTTTACTAGCTTCTGCTGCAATTTG 660
DB 7476 TCTACGCTCTTTGCGTATGAGGCTAATATTTCTTTACTAGCTTCTGCTGCAATTTG 7417
QY 661 TTGGGCTTCAAGATTTGAGAGCTTCCGTTTCTGAAGGATGTAGATCTTTGGGAATTTCTT 720

Db 7416 TTTGGGCTTCAAGGTGAAGCTCCGTTCTGAAGCTAGATCCCTGGGAATTTCTT 7357
 Oy 721 TAGCTTTTGGATGGCTATGATGTATGCTGGACTTCTCTTANNGCAGTACTGCT 780
 Db 7356 TAGCTTTTGGATGGCTATGATGTATGCTGGACTTCTCTTANNGCAGTACTGCT 7297
 Oy 781 GGATCAATAGAAGCTATTGACCGATCCGCTTCTATATCAAGAAATGCAAAAG 840
 Db 7296 GGATCAATAGAAGCTATTGACCGATCCGCTTCTATATCAAGAAATGCAAAAG 7237
 Oy 841 GAAAAAAGGTCTAACTAAATGATATGAAGATAGCTTCTATCTTGTAGAT 900
 Db 7236 GAAAAAAGGTCTAACTAAATGATATGAAGATAGCTTCTATCTTGTAGAT 7177
 Oy 901 CTCCTATATCTTTATTAATCTCTGTTGTTATGCTTATGCTTATGCTTATTA 960
 Db 7176 CTCCTATATCTTTATTAATCTCTGTTGTTATGCTTATGCTTATGCTTATTA 7117
 Oy 961 TCGAAGTACTTGAAGAAGTCACTGAACACTCAATATCTATATATGATGACTATAGT 1020
 Db 7116 TCGAAGTACTTGAAGAAGTCACTGAACACTCAATATCTATATGATGACTATAGT 7057
 Oy 1021 ACTTATGGGGAAGTCTCTCTTCTGAGCTGGCTGATGCTTACTTATCATGCTATTG 1080
 Db 7056 ACTTATGGGGAAGTCTCTCTTCTGAGCTGGCTGATGCTTACTTATCATGCTATTG 6997
 Oy 1081 TTGGTGTAGTCACTGATGTAATTTGATGTTGATTAAGTGGAGCCCTAGTCACTCGTCA 1140
 Db 6996 TTGGTGTAGTCACTGATGTAATTTGATGTTGATTAAGTGGAGCCCTAGTCACTCGTCA 6937
 Oy 1141 TGGTCTCTCTTAACAGTATCGTTTCTGCTCTTCTTATCTTATAGAAACCAAGCTTCG 1200
 Db 6936 TGGTCTCTCTTAACAGTATCGTTTCTGCTCTTCTTATCTTATAGAAACCAAGCTTCG 6877
 Oy 1201 GGCTGGTGGCTATGTTGGGTACAACCTCTCTGATGCTAGCTGGTGTGCGAGCTATAC 1260
 Db 6876 GGCTGGTGGCTATGTTGGGTACAACCTCTCTGATGCTAGCTGGTGTGCGAGCTATAC 6817
 Oy 1261 AGAATATCTCTTCAAGATTCACAAATACGCTCTCTTGAAGTCACTAAGAAATGGCT 1320
 Db 6816 AGAATATCTCTTCAAGATTCACAAATACGCTCTCTTGAAGTCACTAAGAAATGGCT 6757
 Oy 1321 ATATCCCTCTTGAACAGAGCAAAAGTCAAGGTGCTATGATGATGTTGCCG 1380
 Db 6756 ATATCCCTCTTGAACAGAGCAAAAGTCAAGGTGCTATGATGATGTTGCCG 6697
 Oy 1381 CCCGCTTGGGAAATCAGAGGAGCTTATATCCACAAGGTTGCTCGTATCTGTGGA 1440
 Db 6696 CCCGCTTGGGAAATCAGAGGAGCTTATATCCACAAGGTTGCTCGTATCTGTGGA 6637
 Oy 1441 GTATGGAGCTTGAACCCCTTATCTTGAAGTATCTTCTTATCATATGCTATTTGGT 1500
 Db 6636 GTATGGAGCTTGAACCCCTTATCTTGAAGTATCTTCTTATCATATGCTATTTGGT 6577
 Oy 1501 TGGTTTTCGCACTAAGTAAACAACATCTTATGAGCGAGTCTCTTAAAGAACAAG 1560
 Db 6576 TGGTTTTCGCACTAAGTAAACAACATCTTATGAGCGAGTCTCTTAAAGAACAAG 6517
 Oy 1561 AAGTGGCTCAGAAATTCAGTCTCTGCTTCTTCAATAGAGTTGCTTCTTACTTGT 1620
 Db 6516 AAGTGGCTCAGAAATTCAGTCTCTGCTTCTTCAATAGAGTTGCTTCTTACTTGT 6457
 Oy 1621 GATCCCTACCTGCTTTT 1637
 Db 6456 GATCCCTACCTGCTTTT 6440

RESULT 2
 AE001619 11421 bp DNA linear BCT 01-DEC-2000
 LOCUS Chlamydia pneumoniae section 35 of 103 of the complete genome.
 ACCESSION AE001619 AE001363

VERSION AE001619.1 GI:4376620
 KEYWORDS Chlamydia pneumoniae CWL029.
 SOURCE Chlamydia pneumoniae CWL029.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE 1 (bases 1 to 11421)
 AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
 MEDLINE 99206606
 PUBMED 10192388
 REFERENCE 2 (bases 1 to 11421)
 AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
 TITLE Direct Submission
 JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
 FEATURES
 source
 1. 11421
 /organism="Chlamydia pneumoniae CWL029"
 /strain="CWL029"
 /db_xref="taxon:115713"
 complement(95. 1069)
 /gene="y1gb"
 /note="Cpn0346"
 complement(95. 1069)
 /gene="y1gb"
 /codon_start=1
 /transl_table=1
 /product="Integral Membrane Protein"
 /protein_id="AADI8490.1"
 /db_xref="GI:4376621"
 /translation="MALGSPDYGVSPFQSPVSPFSLFSGSLFTSLYIDIOIIVE
 LAISGSAFAGCTFLVRKMYANAVSHHTVFLGVCCLFTHOITSLCTLLAAM
 TAITGFLITRNTFRVSESSALVFLSFLSLVLVLPRTNATIGELVGNAD
 SLTKEDIEPVATITLANAVITTFAPRSLVCSFDSVASSLGIPIRLVDYLIIFQISA
 CLVAFRAVAGLALAFLLIPLSLAKVIASIRSLAMSLVFSIGTAFVLPASSRAIL
 SAYDLGSTGSISVFLTMVYIVKFIYFRGYSKNFERISEKSSQY"
 complement(1069. 2415)
 /gene="y1gc"
 /note="Cpn0347"
 complement(1069. 2415)
 /gene="y1gc"
 /codon_start=1
 /transl_table=1
 /product="Integral Membrane Protein"
 /protein_id="AADI8491.1"
 /db_xref="GI:4376622"
 /translation="MLSCVPSDFIFLSSFLAVTLICMTALMGTILMGTILISKOPLESSE
 SHASYPLVGLALMAQYVFSLOSIFWIVLFGCAVSFGYGIIVFLGKVKLHKDSAL
 CFYVVFPAIGVILASVYKESPLVLRINAYLIGQATLGLEATLAAIVCFSLPA
 LMWYROIIVTTTPDKRAVYCGIKTVYEALESILFISIVSGVRSVGIYLSAMPFA
 PSICARLSDRLSTILLLSAFPGGISALSYISVATCAITIGQAVPTLPGLYV
 VTCAGLILAGLCLLSPKSGWIRFVRKRHFSSKDOEHLKVFPHISHNLNENTSV
 FVCSYKQYQFEGPPEPRMRYOILDMRGYVKKEDDYRLTRKKGSEALRLVRHRLME
 SYLVNSLDSFESKESVHELAEIEHLTEELDHTLTLEILNDCYDPRHOIIPNRKEV"
 complement(2418. 3197)
 /gene="y1gb.1"
 /note="Cpn0348"
 complement(2418. 3197)
 /gene="y1gb.1"
 /codon_start=1
 /transl_table=1
 /product="rRNA methylase"
 /protein_id="AADI8492.1"
 /db_xref="GI:4376623"
 /translation="MNVDKDEFMSVHNLVNYEHAVALYHISFLSGSLTALIGPNG
 AGKSTLLKASLIGLITPSSGTYVYFNQKFKKRVORIAVMPORASVDMFPPTVLDLIM
 GCYSYKGMGRISDDRRERAFHILERGVLSVADROIOLSGSGOORAFILARLIMOKA
 DIVYIMDELFSAIMASFKTSVGVLOELRDGKXTIVYVHHLSHVRQLFDHIVLNLRL
 ICCGPTDECLNGDTIFQTYGCEIELELROTILKLSHGKQFGSC"
 complement(3181. 4164)
 gene

```
gene
CDS
    /gene="ytga"
    /note="Cp0349"
    complement(3181..4164)
    /gene="ytga"
    /codon_start=1
    /transl_table=11
    /product="Solute Protein Binding Family"
    /protein_id="AAD18493.1"
    /db_xref="GI:4376624"
    /translation="MDAKMGYIEKVMRIFCFVACGTFGCTNSGFQANSPCLLS
    NRMTHDCVERVGNRLATVILKSGLDHAEVWKGDOKTAGSAVIFCNLSGHEHTL
    SLRKLNNPNPSVNGERLILARGAFLPEEDGICDPHTMDLSIKWEKAVITEVELE
    KPEWSEAFKANSSELVCEMSILDSMAKOCSTIIPENLRYLVSGHNARSYTRRLAT
    PEVVASGARSRCISPEGLSPEADISVDIAVVDYINEDHVSVPPEPDTLNODALKK
    IVSSLKSHLRYLAKQKPLYSNDVDNVSFTKHNVCCLTEELGVALBCQR"
    4575..5081
    /gene="Cp0350"
    4575..5081
    /gene="Cp0350"
    /codon_start=1
    /transl_table=11
    /product="CT066 hypothetical protein"
    /protein_id="AAD18494.1"
    /db_xref="GI:4376625"
    /translation="MAVEOSHIEKEIEKLGKAIKRVGNKENDLCRYLPSPSGYMH
    HFTLKMSAPBOLLKMLKTEIESEPTIINPPRAPRSGSKRDEINFTKTDIER
    VLRLARQGDMDLARSFKKPLSLKRELLRSIRNGIVSELMNAVYEAUKAVSPN
    LEVNSPY"
    5282..6829
    /gene="adt_1"
    /note="Cp0351"
    5282..6829
    /gene="adt_1"
    /codon_start=1
    /transl_table=11
    /product="ADP/ATP Translocase"
    /protein_id="AAD18495.1"
    /db_xref="GI:4376626"
    /translation="MTREKPKGKLSFLMIPHTHELKKVLPMPLEFCTTNTVL
    RDKDTLVGAPSGAEAFPIKFMVVPCHITIMLYAKLSLISKQALFYAAGTPF
    LIFLAFPTVYIPLEDDVLPTEFADRLQALLPGLGLVALIRNMTFAEYVLELWG
    SVMLSLMGFANEITKTHEAKREFALGICANISLGSRAIYASKLRASVSGVD
    PWGISRLMAMTIVSGLVNMYWINKNVTDPREFNPEDMGKKGAKPKMMKD
    SFELARSPYILLALVIAYGICINLEVMWSQLKQYNNMNDYSEPMGNSFTSG
    VSVYLIMFVGNGVIRKRGMLTGALVPMVNLGIVFEALVIPRNOASCIYAMGCTP
    PLMLAVYGAIGNISKSTKYLDPSTREKAYIPLDQCKYKGAIDVAREKSG
    GALIQGGLVIGSIGAMTPYLAVILFTIAIWLVSATKLNKFLAQSALKEOVAOE
    DSAPAS"
    6968..8227
    /gene="Cp0352"
    6968..8227
    /gene="Cp0352"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18496.1"
    /db_xref="GI:4376627"
    /translation="MKCSPLTVPHIPIKNDCECHRSCLKIRTIARILISVLAYS
    ALSFYFLAARISYIAGTLLAALVILITTVVALAKSKVLPINELQKITNNYRK
    EYFYFVKTHSLVNELEKIFINCMKSGDLPNHLKKAFAFIDILKSIDLTFPEFEE
    ILQNCPLVWLSHEIDKTESVAGEIGLKKTKVYGLGLPLAFHNGYTTIFHSYPLL
    TLISEQKFLYSKASKNQMDSPVYKCEIEFKELPHMLFRKDOVQISQELFELS
    HGITWEOQMOLINPDNMKMLCOPDKAGGCEISMATPCGPIINTEGMDPVSAYEPT
    VPMTKRELKYLLEKVESPMHPSALYOKICVMTTHQNLKRMQVRNYSOWTSS
    LPQAFHQITYLEKRIESSLPINSSL"
    8287..8532
    /gene="Cp0353"
    8287..8532
    /gene="Cp0353"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18497.1"
```

```
gene
CDS
    /db_xref="GI:4376628"
    /translation="MRFNKISLIIFIKIRYSQSGPKQKARPEFKKINSLSVILL
    LEAFENESSSTIQNFKNFKNNKINISIRLPVFTT"
    8561..9904
    /gene="Cp0354"
    8561..9904
    /gene="Cp0354"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18498.1"
    /db_xref="GI:4376629"
    /translation="MSNTSPVIONRSCNYPPELKNSTTHIYISATILGALIAFL
    CVAPVSYILSGALIGLILALIGVILGKIKTPMISKREOVPOELVYNRIRAHYK
    FVSDFSEAKRNKLDLISFTDLLNOLSEVGSSTNYNSELOOKITFESIRANKNE
    VRTASLRLSASASRLPEFLPKTILOKVPFPEWGEETISAGSVVELLHVRKKTGSL
    EEDLSDYIKRPMELPYWLIPLDFRPTNSIINLTHVLARVLARVDFQHLKALNGE
    WNLHSDLNTMKQOLFARXHAAYOSYKHLSPQSDQEDFVNLILCIFKHRYSMQMSL
    IKTVPADLMEMLCITLDHTRGPQDMEFASLIGLYTQGLIHKSEARLSTLISLD
    QFKTRRQSTNIAMFLEMLATHNSFTRSLPITYHPLKRSVFSQPEDESSLIG"
    complement(9982..11283)
    /gene="Cp0355"
    complement(9982..11283)
    /gene="Cp0355"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18499.1"
    /db_xref="GI:4376630"
    /translation="MSMTIVPHALFKNHCESHSTFPLSSRTIIRVLAISFLCIGALAA
    LGCLAPVSYIVSYLAFVILISIVLALIFGKKKAPPTPIIDRFTHIDEAVG
    LSISAFVREOQVITAEFRQESTALLCNISPEEKIKOLPSELRSKVESGIRLADLE
    KNNPFIREDLSQCPLYWLOKFTSAGDPQVCRDGVPRECTGYWLPGLYSTAKAT
    IFCKETHHI1QOILTRKEDVLLKNAALOKWODEKALVERIYTYTYARGLKTEAG
    LTKETISKELLLSLHGYSFDOLILQILPDMADWICFVNSSTAYNLOLCAVGLALS
    SONLDESSIDFVNLGLYVLODLKEAVQASADEPKELGKFLRLHSVSKRLS
    VLROGLIHAIHEGNARARVYDVNPTCARIRHRTSIFFXD"
    BASE COUNT      3791 a      2376 c      1936 g      3318 t
ORIGIN
Query Match      90.7%; Score 1484; DB 1; Length 11421;
Best Local Similarity 99.8%; Pred. No.:0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      1  GAAATAAAAAAGCTATACAGATGCAAAATPAAAGTATTTACAGGGTAAATATGACAAAA 60
      |||||||
Db      5232  GAAATPAAAAAAGCTATACAGATGCAAAATPAAAGTATTTACAGGGTAAATATGACAAAA 5291
QY      61  CCGAAGAAAAACCTTTGGAAATTTGGCGCTCTTCTGTGGCCGATPACTACAGAGC 120
      |||||||
Db      5292  CCGAAGAAAAACCTTTGGAAATTTGGCGCTCTTCTGTGGCCGATPACTACAGAGC 5351
QY      121  TAAAGAAAGTTCTGCCAATGTTCTAATGTTCTTGTGATTAACATTTAATACGAGTGT 180
      |||||||
Db      5352  TAAAGAAAGTTCTGCCAATGTTCTAATGTTCTTGTGATTAACATTTAATACGAGTGT 5411
QY      181  TACGGGATPACAAAAGACATCTTATTGAGGAGCTCGGTCTGGTGTGACAGGCAATTC 240
      |||||||
Db      5412  TACGGGATPACAAAAGACATCTTATTGAGGAGCTCGGTCTGGTGTGACAGGCAATTC 5471
QY      241  CTTTCATCAAGTTTGGCTGTGTTGCCCTGTGCTGATTAATCTTATMGCTATTATTGCAA 300
      |||||||
Db      5472  CTTTCATCAAGTTTGGCTGTGTTGCCCTGTGCTGATTAATCTTATMGCTATTATTGCAA 5531
QY      301  AGCTAAGTAATATTTTAAGTAAGACAGGCCATTTATTATGACGTGGAAAGCCCTTTTAA 360
      |||||||
Db      5532  AGCTAAGTAATATTTTAAGTAAGACAGGCCATTTATTATGACGTGGAAAGCCCTTTTAA 5591
QY      361  TTTTCTTTGCCCTGTTCGCCGAGCTGTAATTTATCCGCTACAGCATGTTTACATCTCAAC 420
      |||||||
Db      5592  TTTTCTTTGCCCTGTTCGCCGAGCTGTAATTTATCCGCTACAGCATGTTTACATCTCAAC 5651
QY      421  AATTGTGTCAGCGTTTACAGGCATCTACCTCCAGGATTCGTAGACATCGTTGCCATCT 480
```

|||||
Db 5652 AATTGCTGACCGTTACAGGCGCTACCTCCAGAGATTGCTAGACTCGTTGCCATCT 5711
Oy 481 TAAGAACTGACATTTCTGATTTTATGTAAGTCTGCACTATAGGGGAAGCCGATGC 540
Db 5712 TAAGAACTGACATTTCTGATTTTATGTAAGTCTGCACTATAGGGGAAGCCGATGC 5771
Oy 541 TATCTATATGTTGGGAGATTGGTAATGAATAATCAAAATCCAGCAAGCAAGCGTT 600
Db 5772 TATCTATATGTTGGGAGATTGGTAATGAATAATCAAAATCCAGCAAGCAAGCGTT 5831
Oy 601 TCTACGCTCTTTTGGGATTCGAGAGCTAATATTTCTTACTAGCTCTGCTGTCATTTG 660
Db 5832 TCTACGCTCTTTTGGGATTCGAGAGCTAATATTTCTTACTAGCTCTGCTGTCATTTG 5891
Oy 661 TTTGGGCTTCAAGTTGAGAGCTTCGTTCTGAGAGTGTAGATCCCTGGGGAATTTCTT 720
Db 5892 TTTGGGCTTCAAGTTGAGAGCTTCGTTCTGAGAGTGTAGATCCCTGGGGAATTTCTT 5951
Oy 721 TACGCTTTTGTAGTGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 5952 TACGCTTTTGTAGTGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 6011
Oy 781 GGATCAATTAAGACGATTTGACCGATCTCGCTCTATATCCAGAGAATGCAAAAG 840
Db 6012 GGATCAATTAAGACGATTTGACCGATCTCGCTCTATATCCAGAGAATGCAAAAG 6071
Oy 841 GGAAGAAAGGCTTAACCTAAATGATGAAGATGAGCTTCCTCATCTGTTATTT 900
Db 6072 GGAAGAAAGGCTTAACCTAAATGATGAAGATGAGCTTCCTCATCTGTTATTT 6131
Oy 901 CTCCTTATATCTTTTATTAACCTCTCTGTTATGCTATGATGATGATGATGATGAT 960
Db 6132 CTCCTTATATCTTTTATTAACCTCTCTGTTATGCTATGATGATGATGATGATGAT 6191
Oy 961 TCGAAGTACTTTGCAAAAGTCACTGAAAGCTGCAATATCTTATGATGATGATGATGAT 1020
Db 6192 TCGAAGTACTTTGCAAAAGTCACTGAAAGCTGCAATATCTTATGATGATGATGATGAT 6251
Oy 1021 AGTTGATGGGAACTTCCTCTGAGTGGGATGATGATGATGATGATGATGATGATGAT 1080
Db 6252 AGTTGATGGGAACTTCCTCTGAGTGGGATGATGATGATGATGATGATGATGATGAT 6311
Oy 1081 TTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 6312 TTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6371
Oy 1141 TGGTTCTCTTAACAGATTCGTTTCTGCTCTGTTATCTTTAGAAACCAAGCTTCTG 1200
Db 6372 TGGTTCTCTTAACAGATTCGTTTCTGCTCTGTTATCTTTAGAAACCAAGCTTCTG 6431
Oy 1201 GAGTGTGCTATGTTGGTGAACACTCTCATGCTAGTGTGTTGGAGCTATATC 1260
Db 6432 GAGTGTGCTATGTTGGTGAACACTCTCATGCTAGTGTGTTGGAGCTATATC 6491
Oy 1261 AGAATATTTCTTGAATAATCAACAAATAGCGTCTCTTGTAGCTCAACTAAGAATAGCGCT 1320
Db 6492 AGAATATTTCTTGAATAATCAACAAATAGCGTCTCTTGTAGCTCAACTAAGAATAGCGCT 6551
Oy 1321 ATATCCCTCTTGACCAAGCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAG 1380
Db 6552 ATATCCCTCTTGACCAAGCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAG 6611
Oy 1381 CCCGCTTGGGAAATCAGAGAGGCTTTAATCCAAAGGTTGGTGTATCTGTGGA 1440
Db 6612 CCCGCTTGGGAAATCAGAGAGGCTTTAATCCAAAGGTTGGTGTATCTGTGGA 6671
Oy 1441 GTATTGAGCTATGACCCCTTATCTTGCATGATTTCTTTTCATCTATTTGCTATTTGCT 1500
Db 6672 GTATTGAGCTATGACCCCTTATCTTGCATGATTTCTTTTCATCTATTTGCTATTTGCT 6731
Oy 1501 TGGTTTCTGCAACTAAGTTAAACAACATATTTCTTACGCGAGCTCTCTTTAAGAACAAG 1560
|||||

Db 6732 TGGTTTCTGCAACTAAGTTAAACAACATATTTCTTACGCGAGCTCTGCTTTAAGAACAAG 6791
Oy 1561 AAGTGCTCAAGAGATTCAGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1620
Db 6792 AAGTGCTCAAGAGATTCAGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 6851
Oy 1621 GATCCCTACCTGCTTTT 1637
Db 6852 GATCCCTACCTGCTTTT 6868
RESULT 3
AP002546
LOCUS
DEFINITION
Chlamydia pneumoniae J138 genomic DNA, complete sequence,
section 2/4
ACCESSION
AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801
AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808
AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815
AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA000008
AP002546.2 GI:10176692
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On or before Sep 15, 2000 this sequence version replaced
gi:6172290, gi:6172292, gi:6172294, gi:6172296, gi:6172326,
gi:6172328, gi:6172330, gi:6172332, gi:6172334, gi:6172336,
gi:6172338, gi:6172382, gi:6172384, gi:6172386, gi:6172388,
gi:6172390, gi:6172392, gi:6172394, gi:6174666, gi:6174668,
gi:6635174, gi:6635176, gi:6635178, gi:6635180, gi:8547433,
gi:8547438, gi:8978640.
Location/Qualifiers
1. 300650
/organism="Chlamydia pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
144. 935
/gene="CPj0267"
144. 935
/gene="CPj0267"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BA08477.1"
/db_xref="GI:8978641"
/translation="MSIMSLNKNALLNPPEAVCLNAMPKXYINDRKTFCTVTL

```

gene
CDS
    957..1343
    /gene="CpJ0268"
    /gene="CpJ0268"
    957..1343
    /gene="CpJ0268"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAA98478.1"
    /db_xref="GI:8978642"
    /translation="MSOCSSSTWEMKSEFVPMKNPTPLSPIPSEDEFILAYEP
    FVLKTDENAOANPGTSTFNVENGIDDLNPLLGQPMEOANNANPGTSSGNPTSLPA
    PERLPETENQEEEGSQNNEDLIG"
    complement(1495..2472)
    /gene="CpJ0269"
    complement(1495..2472)
    /gene="CpJ0269"
    /codon_start=1
    /transl_table=11
    /product="olipeptidase"
    /protein_id="BAA98479.1"
    /db_xref="GI:8978643"
    /translation="MTIDHCDLISHPHCRKDPVRCSPOLLSCGVRQOVCATFVP
    HSRSEPCNDKONSLEFSLPNQYPRDIGLSYEEREGSSQKSLSLSTENASALGD
    DTPALGTLLAKLILHTGQGPLAYLGIVKNGNRGQTEAPKRLSNDKVLDDIYEL
    GVPIDLSHCSDKLAEIDIDYADKLPNLVLAISHNFSVLDHRNLDAHAKEIVRR
    KGVIGLNLVRSYVGDLSLDLEKHVLAENGLISLIVSGDFYANEDENEFNCCSS
    AEAHPVNLQHLHRFESKKAESILSSRAEKLKQYIVQVMPKITDVKL"
    complement(2487..3347)
    /gene="ywlC"
    complement(2487..3347)
    /gene="ywlC"
    /codon_start=1
    /transl_table=11
    /product="Suas superfamily-related protein"
    /protein_id="BAA98480.1"
    /db_xref="GI:8978644"
    /translation="MPDKKAQITFSLPVMSAIIHQKIVALPTDTVYGVFLVYASEA
    EERLYALKDREPSKAFALYVNSIEDIENISGYPSPTAKKLAQLFPGAITLVVHRNP
    RFPKETTAFRIVDHSVREIVDHCGTILIGTSANLSEPSALTAQRIEADPFADHLCIF
    DPGCSHGLESTIVASDPLIYIREGLISRSVLEINAGTEAKIFHRTSHAFSKHIIKYTV
    KNOBOIVSFUSGSLDFKGVCEHPRKPFYTRLRKALKKRTPSYIYIDINTSDPEL
    PFLSPYIYE"
    3656..4387
    /gene="CpJ0271"
    3656..4387
    /gene="CpJ0271"
    /codon_start=1
    /transl_table=11
    /product="lysophospholipase"
    /protein_id="BAA98481.1"
    /db_xref="GI:8978645"
    /translation="MTDYSFPRKLTGNTEALTECRNPDDPIITLICHGGSALDNTFF
    PSICSFSLRPWTWIFPNKGIPLDENDFRGSRACPLNVLLOELSLYANGCNLOEKY
    DELEDVDLETEREALEELLINLRPYNETIIGGFSQGAIIATHVLVTSQNPYAGALIL
    AGARLFGNGWEGEELKQCAQVPELOSHGDELDLPHLGHANLDLLTLKNGCFVSFHG
    GHEIPSYVFOKQVTVPMWIDPARG"
    complement(4359..5255)
    /gene="dnaX_2"
    complement(4359..5255)
    /gene="dnaX_2"
    complement(4359..5255)
    /gene="dnaX_2"
    /codon_start=1
    /transl_table=11
    /product="DNA polymerase III gamma and tau"
    /protein_id="BAA98482.1"
    /db_xref="GI:8978646"
    /translation="WHLDENGWEALIRKYVHOEVPPAILLHGFLLFVLDOKAEOLA
    SEILSSPSGESEHKVOKIHPDIYOFPEPGGRSLSIDLPKIKQIYISFEAVYKI
    YIIHADRMTLAISAFLKVEEPKHAVALIITLTKVQLRKPRTIISRLSIFIEGEEK
    ILCSEKTSYLFRAQCEIPIVTEVSQIIKESSETDKVLRDQVGRFEMVLLLELYDRY

```

```

gene
CDS
    complement(5246..5866)
    /gene="tdk"
    complement(5246..5866)
    /gene="tdk"
    /codon_start=1
    /transl_table=11
    /product="thymidylate kinase"
    /protein_id="BAA98483.1"
    /db_xref="GI:8978647"
    /translation="MTFIVEEGESGKSLAKALGDQVLAODRKVLLTRREGCLIGE
    RLRLDILEPPLLESCCELFLFGSRQHOEYVIPALRGDYIVICERFSDTIVYG
    GIAREGADVFADICSKVSGPTPLPNVILLDIPADIGLQRKHOKVQEDKEKKPLS
    YHNRIREFLSLASDPSRYIVLDAKRESLADIKVMLHTQLGCT"
    complement(5871..8375)
    /gene="gyrA_1"
    complement(5871..8375)
    /gene="gyrA_1"
    /codon_start=1
    /transl_table=11
    /product="DNA gyrase subunit A"
    /protein_id="BAA98484.1"
    /db_xref="GI:8978648"
    /translation="MPNKDEIIVRNLEEMKESLYRSMSVITSRALPDIRDGKPS
    ORVLYAMKQI.SLSPGAKHRCACIKGTSDDYRPHGSESVIYPTLVRAQWAMARPL
    VDGQNGFSDIGDPPRAARVYEARTLTHSAMLMEDLDQVDIVDPNDETHERVPEP
    SKFPLNCGSSGGLAVGMATYIPHNILGELATLLANQASVDELQWPGDFP
    TGGICGEGEGRSTYTGRRGKIKVARLHVENEDKHRESITITEMPVNYSKRLIEG
    IANLYNENKTISDVREDSQGRIVALEIKGSSSTIINRLYKFPDVQVTFGAM
    LADKNLPTMSIHMTISAMTIRHREKVRIRRTREYNKAETRAHYLEGYIKALCGLDA
    LVKTIRESGNKEHAKERTIESFGPTREPALKALBELRTQLGLEPEKIOKYEELINK
    IAYTKQVLSDEGLVKDITRNELDLKHKKVARRTTEFDADDIRDIEDIITNESVIL
    TISGDVYKRMVAVYFKEQRRGGGVGTFDMKKGAGFLKAVYSATKCYLLIFNFGQ
    CYWLKVMQLPGEERRAKGPIINFLGIRPEBELAAILININPNAGFLPLATRGVY
    KKVSLDASNPBKKGIGIRALEIDEGBELIAACHIVSDEEKVMLFTHLGMAVREPEKVR
    PMGRARGVRCVSLKNEEDKYVSCQIYENOSVILVCDQGGKRSIVDERFETRGCV
    GVRSLINERGNVIGATIPVTDHDSILMSSQGAIRINMDVYVMGKSTOGVRLVHL
    KEGDALVMEKLSNENDEVLSGSEECSDTVLSR"
    complement(8391..10808)
    /gene="gyrB_1"
    complement(8391..10808)
    /gene="gyrB_1"
    /codon_start=1
    /transl_table=11
    /product="DNA gyrase subunit B"
    /protein_id="BAA98485.1"
    /db_xref="GI:8978649"
    /translation="MDPEKKNYDASAITVLGLOAVRPMYIGDNGITGLHLVYE
    VVNDSIDPAMGYCSRIDVRILEDQGIYIVNGKCIPEVHERESAKOGRVSALEY
    LTVLHAGKEDFDSTKYSGGLHGVGCVNALSEKLVAIVFKDKKCTOMERSKGIPTV
    PLQVYVSDRQGTETVEYDPDKIFSTCFEDSILMKRLRELAFINRGITVLEDDRQV
    SFDKVTFPEBGIOISFYSLNQNKSELSSEPIYICGTVGDDGEIEFEALQNMNGYS
    ELVYSYANNIPTRQGGTHLGEFTALPTVIMYIKAHNLAKNNKALKALGEOIRGLTA
    VLSKVPMPQFQEGTKOKLIGNSDVSAVOQVGEALTTFEPENQIAMIYIDKVFVA
    QAREAKARBELTLKRSALDSBARLPKGLIDLENDPECEKNIYBGDSAGSAAQGRD
    RRFQATLPTKGLTLNVEKARLQKIFQNEIGTIIAALCGGAGDAPFNLSKRYRIIT
    MTDAVDQSHIRTLITFVYRHMTALINECVYIAQPLVYSSKKERYLLISEKEMD
    SYLLMAGNESISILFKSTERYELGALSEFINVLIVDSFINTLEKKAIPSEFLEWY
    KEGIGYPLVYLAAPQMGQGRIVYSDKEEKALQOETHRKITEILYVVAAYEIQIONG
    LKEYGDISYVLIPOKNEIYVGINEDSPSCNYSCTLEVIYINKLNGKGIETQRYNG
    LGEMADQLMDPTTANPQORTLIHVSILKDAVEADHITFLMKGEVPPREFTESHALSI
    RINNLDT"
    complement(10812..11159)
    /gene="CpJ0276"
    complement(10812..11159)
    /gene="CpJ0276"
    /gene="CpJ0276"

```

```

Query Match 90.7%; Score 1484; DB 1; Length 300650;
Best local Similarity 99.8%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAAATTAATAAACGATTCAGAAATGAAGTATTTCAGAGCGTAATATGACAAAAA 60

```

```

|||||
Db 93946 GAAATTAATAAATATCAGATGAGAAATAAAGTATTTCAGAGGGGAAATATGACAAAAAA 94005
Oy 61 CCGAAGAAAACCTTTTGGAAAATTCGGCTCTTCTTGTGGCGATACATACACAGAGC 120
Db 94006 CCGAAGAAAACCTTTTGGAAAATTCGGCTCTTCTTGTGGCGATACATACACAGAGC 94065
Oy 121 TAAAGAAATCTTGGCCAAATGTTCTTAATGTTCTTCTGATTAACATTTAACTATACGGTGT 180
Db 94066 TAAAGAAATCTTGGCCAAATGTTCTTAATGTTCTTCTGATTAACATTTAACTATACGGTGT 94125
Oy 181 TACGGGATACAAAAGACACTCTTATTTGTGGAGCTCCTGGTCTGGTGCAGAGCAATAC 240
Db 94126 TACGGGATACAAAAGACACTCTTATTTGTGGAGCTCCTGGTCTGGTGCAGAGCAATAC 94185
Oy 241 CTTTCAATCAAGTTTGGCTGTGTTGCCCTGAGCTATTAATCTTTATAGCTTATTTATGAA 300
Db 94186 CTTTCAATCAAGTTTGGCTGTGTTGCCCTGAGCTATTAATCTTTATAGCTTATTTATGAA 94245
Oy 301 ACCTAAGTAATATTTTAACTAAGACAGGCTTATTTTATGACAGTGGAGACGCCCTTTTAA 360
Db 94246 ACCTAAGTAATATTTTAACTAAGACAGGCTTATTTTATGACAGTGGAGACGCCCTTTTAA 94305
Oy 361 TTTTCTTGGCCCTGTTCCCGACTGTAATTTATCCGCTAGCGCATGTTTACATCCCTACAG 420
Db 94306 TTTTCTTGGCCCTGTTCCCGACTGTAATTTATCCGCTAGCGCATGTTTACATCCCTACAG 94365
Oy 421 AATTTGCTGACCGTTTACAGGCCATTCCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 480
Db 94366 AATTTGCTGACCGTTTACAGGCCATTCCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 94425
Oy 481 TAAAGAACTGGACATTTTCTGCAATTTTATGTAATCTTGTCTAATATGAGAGCGTCATGC 540
Db 94426 TAAAGAACTGGACATTTTCTGCAATTTTATGTAATCTTGTCTAATATGAGAGCGTCATGC 94485
Oy 541 TATCTCTAATGTTTGGGATTTGCTAATGAATTAACAAATTCACAGAGCAAGAGCGCT 600
Db 94486 TATCTCTAATGTTTGGGATTTGCTAATGAATTAACAAATTCACAGAGCAAGAGCGCT 94545
Oy 601 TCTAGCGCTTTTTCGATTCGAGGCTAATATTTCTTTACTAGCTCTGTCGTCGAATTC 660
Db 94546 TCTAGCGCTTTTTCGATTCGAGGCTAATATTTCTTTACTAGCTCTGTCGTCGAATTC 94605
Oy 661 TTTGGGCTTCAAGTTGAGAGCTTCCGTTTCTGAAGGTGTAGATCTTGGGCAATTTCTT 720
Db 94606 TTTGGGCTTCAAGTTGAGAGCTTCCGTTTCTGAAGGTGTAGATCTTGGGCAATTTCTT 94665
Oy 721 TACGCTTTTGTAGGCTATGACTATTTGATCTGACCTGTTCTTATGAGCAATTTACTGCT 780
Db 94666 TACGCTTTTGTAGGCTATGACTATTTGATCTGACCTGTTCTTATGAGCAATTTACTGCT 94725
Oy 781 GGATCAATAGAACCTATTTGACCGATTCCTCGCTTCTATATCCAGAGAAATGCAAAAG 840
Db 94726 GGATCAATAGAACCTATTTGACCGATTCCTCGCTTCTATATCCAGAGAAATGCAAAAG 94785
Oy 841 GGAAGAAAGGTGCTTAACCTAATAATGATGAAAGATAGCTTCTCTATCTTGTATAGAT 900
Db 94786 GGAAGAAAGGTGCTTAACCTAATAATGATGAAAGATAGCTTCTCTATCTTGTATAGAT 94845
Oy 901 CTCCTTATATCTTTTATTAATCTCTGTTGATGCTATGCTATGCTATTTGCAATTAATCA 960
Db 94846 CTCCTTATATCTTTTATTAATCTCTGTTGATGCTATGCTATGCTATTTGCAATTAATCA 94905
Oy 961 TCGAAGTACTTGGAAAAGTACAGTGAACATGCAATATCTCAATATGATGATGACTATAGC 1020
Db 94906 TCGAAGTACTTGGAAAAGTACAGTGAACATGCAATATCTCAATATGATGATGACTATAGC 94965
Oy 1021 AGTTATGAGGACCTTCTCTTCTGAGCTGGCGTATGATTCGCTACTATTCATGCTATTTG 1080
Db 94966 AGTTATGAGGACCTTCTCTTCTGAGCTGGCGTATGATTCGCTACTATTCATGCTATTTG 95025
Oy 1081 TTGGGTGTAAGTCAATGCTAATTTGATGATGTTAAGTGAAGCCCTACTCACTCTGTGCA 1140
|||||

```

```

Db 95026 TTGGGTGTAAGTCAATGCTAATTTGATGATGTTAAGTGAAGCCCTAGTCACTCTGTCA 95085
Oy 1141 TGGTCTCTCTAACAGGATATCGTTTCTTCTGCTGTTATCTTTTGAACCAAGCTTCTG 1200
Db 95086 TGGTCTCTCTAACAGGATATCGTTTCTTCTGCTGTTATCTTTTGAACCAAGCTTCTG 95145
Oy 1201 GCGTGGTGGCTATGTTGGTGTACACTCTCTCATGCTATGCTATGCTGTGGAGCTATAC 1260
Db 95146 GCGTGGTGGCTATGTTGGTGTACACTCTCTCATGCTATGCTGTGGAGCTATAC 95205
Oy 1261 AGAATATCTTCTGCAAAATTCACAAAATAGCTCTCTTTGACATCACTAAAGAAATGGCT 1320
Db 95206 AGAATATCTTCTGCAAAATTCACAAAATAGCTCTCTTTGACATCACTAAAGAAATGGCT 95265
Oy 1321 ATATCCCTCTTACCAAGAGCAAAAAGTCAAAAGTAAAGCTCTATGATGATGTCGCG 1380
Db 95266 ATATCCCTCTTACCAAGAGCAAAAAGTCAAAAGTAAAGCTCTATGATGATGTCGCG 95325
Oy 1381 CCCGCTTGGAAAATCAGAGAGAGCTTTAATCAACAAGTTTGTCTCTATCTGTGAA 1440
Db 95326 CCCGCTTGGAAAATCAGAGAGAGCTTTAATCAACAAGTTTGTCTCTATCTGTGAA 95385
Oy 1441 GATATGAGCTATGACCCCTTATCTGCAAGTATCTTCTTCAATGCTATTTGGT 1500
Db 95386 GATATGAGCTATGACCCCTTATCTGCAAGTATCTTCTTCAATGCTATTTGGT 95445
Oy 1501 TGGTCTTGGCAACTAAGTTAAACAACACTATTTGAGCCAGTCTCTCTTAAAGAACAG 1560
Db 95446 TGGTCTTGGCAACTAAGTTAAACAACACTATTTGAGCCAGTCTCTCTTAAAGAACAG 95505
Oy 1561 AAGTGGCTCAGAACATTCAGTCTCTGCTTCTTCATATGATTTGCTTACTCTTGT 1620
Db 95506 AAGTGGCTCAGAACATTCAGTCTCTGCTTCTTCATATGATTTGCTTACTCTTGT 95565
Oy 1621 GATCCCTACCTGCTT 1637
Db 95566 GATCCCTACCTGCTT 95582

```

RESULT 4
AE002301/c
LOCUS 23773 bp DNA linear BCT 26-MAY-2000
DEFINITION Chlamydia muridarum, section 33 of 85 of the complete genome.
ACCESSION AE002301 AE002302 AE002160
VERSION AE002301.2 GI:8163203
KEYWORDS
SOURCE
ORGANISM Chlamydia muridarum.
Chlamydia muridarum
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
1 (bases 1 to 23773)
 Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
 White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
 Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
 Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
 Genome Sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935
REFERENCE 2 (bases 1 to 23773)
 Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
 White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
 Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
 Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
COMMENT On or before Jun 1, 2000 this sequence version replaced gi:7190382,
 gi:7190372.
FEATURES
source Location/Qualifiers
 1..23773


```

gene
CDS
/organism="Chlamydia muridarum"
/db_xref="taxon:83560"
/!note="synonym: Chlamydia trachomatis Mopn"
106..1842
/gene="TC0330"
106..1842
/!note="protein export protein, FHIPEP family: identified by match to Pfam protein family HMM PF00771"
/codon_start=1
/transl_table=11
/product="protein export protein, FHIPEP family"
/protein_id="AAF3547.1"
/db_xref="GI:8163204"
/translation="MNHTRGSGMKVTPTPICLTPVLPDIPRMTIDPCVCANLACSTSIIFWFSLRSSASARIPPSLLIYICLLGLNLASTRNITLASGMASPLIFALGFEFSLGSIPALVTLCELEDFINFVITKGAEIAEVARPSTLESIPKQMSLDADIISAGNIGSRASVKNKSLLESDYFSAMEGVFRPVGDAIMSVLLGVNSIALPFGRIHGVNGLLTVLDALASQVPALITSCSAATLAKGEKESLAQHLVEYEOCRQSFYIALITLGMMAFIPGARKALITGFSVLLIGYKSPSEPSILPKERITELLIPKGMENPVLVRAARHOIYOEGLVFPPESSVYRVYNNVSPRVFSQGEISITELSCANLAKRKIAPETIGERTIARLFEBOGGSLIEITIPKTSNSLVPLRALVKEQVSLHPKILLEXIDLFGSQKSSLEIVESRYNLGKQIGLSMNRKDVLEIVDSLVEQFVRDSQAKVFVDLNEKVSQVKDLIREGEGNFRATVGTSETRKEIKRIVDPFPLDLVLARSSELPEEMRITLLGAVSDEVLS"
1950..2711
/gene="TC0331"
1950..2711
/!note="similar to GB:X68709 PID:47122; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor, sigma-70 family"
/protein_id="AAF39194.1"
/db_xref="GI:7190373"
/translation="MKTODLADTMQLYWSTKEIHHRDLFIAYLPVKNVNAHRLASGMPSHVKIEDLVASGVGLYRAVERPDEPKSPESYALPIIKAAITDDLRKQDWPRSVYORANRLADAMDAALBOTLCKEPTGDCLEYNISQELSHMFSSSRPALVSLNDPVS CODEDEGALTEERLADERANGYDVTIRKKEAISITELSLSLDEREKERVMAIYYDDLVLKERIGTLGVSESRVQSQIHSKALIKLGTLSLL"
2869..4107
/gene="TC0332"
2869..4107
/!note="similar to SP:P00952; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="tyrosyl-tRNA synthetase"
/protein_id="AAF39195.1"
/db_xref="GI:7190374"
/translation="WQQLIDSLQKRGILDNSSAGLESITAPYSAYLGFDPAAPSIAHIGHWIGICFLRRLSAYGITPILAVGATMGIDPSGVSERSLDQOVLDNSKKEIALANYLPDIRIVNNAWLGSLMWDLFADIGKYLFGSLMAKDVAKORVSEEGISTYTERSYLLLOSYPHALFKFHGCVLQCGSDQMGNTSSIDYIRKRGAGARGLYPLITLDSKGKTKGSGESVWMLDPELTSPYELFOYFLRLSDOEPIKIAKMLITDDDEVLADKRLNDPQAVKRKYAAEYIVADYHAGBCLAQALATTESPFANKGNITSESLAALVQSGVGINVARADYIGKRWLDVVOLGFCSSKGARRLIQRGLYVNOEPLIDBOSVLDGYTICFDRYTLISQGRKKKQVIDLN"
4117..5556
/gene="TC0333"
4117..5556
/!note="similar to SP:P00349 PID:1193; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="6-phosphogluconate dehydrogenase, decarboxylating"
/protein_id="AAF39196.1"
/db_xref="GI:7190375"
/translation="MAPADIGLIGLAVMGKNLVLMNIDHGAFAVSYNRSPEKTEELFK

```

```

gene
CDS
EHCENISLQGFATIEEFVQSLKRPKIMIMIKAGAPVDEMISLLPFLBEGDILIDGNSYVLDSESRVYIDLKKKGLIFVGMGVSGEGGARGSPITMGNGIEAMPVLAIPFOSIAAONDGQCCSCSTIGGAGHGVKAVAHNKIEVGDIDLICETTEIILKSRDLSEIOIGNIFEFMNQTDLNSITLWGSAAVHLAKDENGVAAASVLLDVAGQKGTGRVADAETAKVAGMSLIIEVLAARYLSAMKEVRQAAKEPVASLLVQPSQEAASVLIEDAREALYAAKIIISYAOGFMILKQIIEERNMDNLNGELALLVIRGCCIIQSAFLDKIHOGFESCPAHSIMLQDYKRVNLVSETGFRFARLTHAVGACVAALPCLASALAFVQVRTENSPFLVQGLRDYFGAHGYERODRPRGFEYTHDMLGSKNASM"
complement(5624..7432)
/gene="TC0334"
complement(5624..7432)
/!note="similar to GB:D17650 SP:P37949 PID:1122398 PID:436036 PID:1303804; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="GTP-binding protein LepA"
/protein_id="AAF39197.1"
/db_xref="GI:7190376"
/translation="MKPYKTENIRNFSTIAHIDHGSTIADRLLESTSTIOEREMQELDSMDLEREGGITKAPVMTYEEEGEYELNLDIPGHVDSYEVSRSLACEGA LLIVDAQGVQASLANAYLALERDLEIIPVNLKIDLPAAQPEAIKKOIEEFIGIDTSLNACSAKTGGGIPETLESITRLVPPPKPSQETELKALIFPSHDPYVGIMVYAVISGEIKGDRITFMSTGSSSEFVLGICAFPEATLIGSLRAGOVGFANIKKVDVKIGDYTVTVKHPAKPELECKETIKPVYFAATYPIIDSSDEPTLDALGRLOLNSALITBEOESSHSLGFGRFCGFLGLHLEITERISREDDLIITAPSVITKVVILKNKNTLETIDNPTAYPDPALEIHEMEPPWVNVNITPOEELSNMSKCMKRGICUKTMDLDQHRLVLSLPLNEIYSDNNDLKYSTKYGSGFDYLGDKPCGAILKLEIILDETVDAFSCIVHRDKAESGRSTCEKLVDPVLPOLFKIPIDAAINKRKIIAETIRALAKNVTAICYGDITRRKRIMKOKKKKKMKKEFGVSIPTNFAVEVLKME"
complement(7617..9206)
/gene="TC0335"
complement(7617..9206)
/!note="similar to GB:M28816 SP:P19568 PID:152470 GB:A135269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ADP, ATP carrier protein"
/protein_id="AAF39198.1"
/db_xref="GI:7190377"
/translation="MTOFAEKPEGKLRSFLMPIHHELKVLPMLFMECISFNVITLBDTKDITVTAPGSAEALPIPIKIMLVYPSVAVFELITAKISNLISKALFYVLSLSPVTFRLPFLVITYPIRHIHPDADTDALIPSGLGTIAMLRNMTFAEYVLSLWGSVMISLMFWGFANETTKISEAKREFALGVAANVALLISGPAIWSKLRASLSEGVDPGWVILFLMAMFETCSCAIIAAGCWMNMRYVLTDPREFNPALAKKSKPKSMGSEFSYLRSPLYMLLALLVICYGVCINVEYTKMSQKMOFPNPNEYSAFPGTSPMTGVVSVFVMLPTGGNVITRRFGMLTGALVTPVVALGATFEFALVIFRDAATGLVAAIGTTPMLAAVYGAONLTSKTYKALFDATKEKAYIPDDEQVYKKAADVYAAARFGSGGSLIQQGLVYCGSISAMTFPLVALVALLIMVLTSKTKLKLPLAASAKBEOLEAVATAAERKASPAKAEVSPAIEGVS"
complement(9553..10029)
/gene="TC0336"
complement(9553..10029)
/!note="conserved hypothetical protein; identified by glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39199.1"
/db_xref="GI:7190378"
/translation="WATAOITTOEFETOLITRAISKVSGSKENDLCRYLPGPGGYMHFTTKRMHSAPEQLKMLKTFVIDSMTPRSINPAPRARGSKRKDDVNTKVDIERVLEIARQVGDHDAFLRFPKRPPLSLKRELIRSRNNIVSTELMNAVYEAALSOPT"
complement(10169..10270)
/gene="TC0337"
complement(10169..10270)
/!note="hypothetical protein; identified by glimmer2; putative"
/codon_start=1

```

```

gene
CDS
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAF39200.1"
    /db_xref="GI:7190379"
    /translation="MOTNENNLKHFIFCCPNQKISKMRLLKKKKPL"
    10658..11638
    /gene="TC0338"
    10658..11638
    /note="similar to PID:790546 PID:1777933 GB:AE000520;
    identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="ABC transporter, periplasmic substrate-binding
    protein, putative"
    /protein_id="AAF39201.1"
    /db_xref="GI:7190380"
    /translation="MEFLHVRKKYKIVIGILCLACGFYISSSSGRKKSIDERHITL
    SMNRITVCVRITGDRKRNLYLDGSDIPHSYEMVKGEDMDMAISQILFCGLGLEH
    SASLRKHLGNSKVIDLGARLDKNCFLVLEDPPDHIMDMGWISYKEMASVL
    VQIIPVAAEEQKNAEQILSEMEDLDRAVSLAIPEKNRYLYTGHNAFSYFTRRYL
    SSDEERESGNMKLRCSPEGLSPEAQISRIDMRVEXICANDVGVFLEDTLNDAL

Query Match
Best Local Similarity 100.0%; Pred. No. 3.5e-13; Length 23773;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 GAGCTAAAGAAAGTCTGCCAATGTTCTTAATGTTCTGTGATT 161
DB 9140 GAGCTAAAGAAAGTCTGCCAATGTTCTTAATGTTCTGTGATT 9096

RESULT 5
TAJ10586 1587 bp DNA linear BCT 23-MAR-1999
LOCUS Chlamydia trachomatis npt1 gene.
ACCESSION AJ010586
VERSION AJ010586.1 GI:4538859
KEYWORDS npt1 gene; nucleoside triphosphate transport protein 1.
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis.
REFERENCE
    1 (bases 1 to 1587)
    Tjaden,J., Winkler,H.H., Schwoppe,C., Van Der Laan,M., Mohlmann,T.
    and Neuhaus,H.E.
    Two nucleotide transport proteins in Chlamydia trachomatis, one for
    net nucleoside triphosphate uptake and the other for transport of
    energy
JOURNAL J. Bacteriol. 181 (4), 1196-1202 (1999)
MEDLINE 99138740
PUBMED 9973346
AUTHORS Neuhaus E.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-1998) Neuhaus E., University of Osnabrueck,
Biology / Plant Physiology, Barbarastr.11, D-49069 Osnabrueck, FRG
FEATURES
    source
        /organism="Chlamydia trachomatis"
        /db_xref="taxon:813"
        1..1587
    gene
        /gene="npt1"
        1..1587
    CDS
        /gene="npt1"
        /function="ATP/ADP transport protein"
        /evidence=experimental
        /transl_table=11
        /product="nucleoside triphosphate transport protein 1"
        /protein_id="CA839534.1"
        /db_xref="GI:4538860"
        /translation="MTOTAEKPFCKLNSFLPFIHMHCLKVLPFLLMFCISNYTTL
        RDKDTLVITAPGSGAEAIPIKILMLVPSAVVMLIYAKLSNLINKOALFFAVLSPF
```

```

VFPALEPVVYPCRHLIHPAFADTLQSLSPGFMFIAMLRMTFAVYVLSLWG
SVMLSLMPFGANETTKISEAKREFYALFGVANYALLISGPAIIWSSKLRASLGEVD
PMGYSILMPFAMFLCSCAIIAACYWMMNRVYLTDPREFYPALEAKSKPKPMGSEF
SYLRSPPYMLLALLIVICGICINIVEYTKMSOLKMPBNNDVSAPMGNSFPMYGVY
SVFWALETGVNITRSTKRYALPDATKEIAYIPLOEQCKRKKAIDVVAAREFGSGGS
MLAVVGAIONILSRSTKRYALPDATKEIAYIPLOEQCKRKKAIDVVAAREFGSGGS
LIQGLLVGCGISGAMPFPLAVLFAIIMWLTSAITKLNLKFLAASAKEDOLEBAAM
AEKSSSAKESAPAEIGVS"
BASE COUNT 384 a 357 c 330 g 516 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 8.4e-10; Length 1587;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 AAGAAAGTCTGCCAATGTTCTTAATGTTCTGTGATT 161
DB 73 AAGAAAGTCTGCCAATGTTCTTAATGTTCTGTGATT 111

RESULT 6
AE001281/c 14168 bp DNA linear BCT 30-OCT-2000
LOCUS Chlamydia trachomatis section 8 of 87 of the complete genome.
ACCESSION AE001281 AE001273
VERSION AE001281.1 GI:3328454
KEYWORDS
    Chlamydia trachomatis.
    Chlamydia trachomatis.
    Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
    1 (bases 1 to 14168)
    Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
    Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
    Koonin,E.V. and Davis,R.W.
    Genome sequence of an obligate intracellular pathogen of humans:
    Chlamydia trachomatis
JOURNAL Science 282 (5389), 754-759 (1998)
MEDLINE 9900809
PUBMED 9784136
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
    source
        /organism="Chlamydia trachomatis"
        /strain="D/UW-3/CX"
        /db_xref="taxon:813"
        86..847
    gene
        /gene="flaA"
        /note="CT061"
        86..847
    CDS
        /gene="flaA"
        /codon_start=1
        /transl_table=11
        /product="Sigma-28/WhiG Family"
        /protein_id="AAC67652.1"
        /db_xref="GI:3328455"
        /translation="MKTHDLADIMQYLSSTKEIHRDVLIESYLVKNAVRLASGM
        PSYHKMEDYALCYVEGLIRAVRRDPKRSKRSFYSALFIITKATITGLRKQMDPPRSY
        YORANRLADMDISRTQTLKEPTDGLCETLNISSQELSHWFSRRPALVLSLHDFS
        CQDDDESLALEERADERADTGVYIRKKAISILITVALLALDEKERQYVVALYDDVL
        VLKEIGKILGVSSRSVSIHSKALKLRKLTLSLL"
```

gene 1006. .2244
/gene="tyrs"
/note="CT062"
1006. .2244
/gene="tyrs"
/codon_start=1
/transl_table=11
/product="tyrosyl tRNA Synthetase"
/protein_id="AAC67653.1"
/db_xref="GI:3328456"
/translation="WQQLIDNLKRRGILIDNSSAGIESLTPVSAVILGHDPAPSLHIG
HMTGICFLRLAAYGITPVALVGATGMIIDPSSKSVSRSLDQAOVIDNSKIKAAAL
ASYLPGIRIVNNAWDLGSLSMWDFLRDYKHFRLGSMALKAVKORVSEGGIIVTER
SYLLQSYDPAHLFEKHNVVLOCSSDQMGNTSGIDYIRRRGGAOAVGLPYLLTDS
KGGIKIGTESGTHWLDPAITPPYELFOYFLPLPOETISKVMRTLLDNEETPALDER
LTPDQAVKAKTIAEIVYADVHSEGLAOQAQATSEFPKSGSLTEALVALVESGVG
VKVARADLIGKRWLDIVELGFCSSRGQARRLLIQGRGLYINOEPLADROSILDTQIC
FDRYVLSQGRKKQVIDLN"
2254. .3696
/gene="gnd"
/note="CT063"
2254. .3696
/gene="gnd"
/codon_start=1
/transl_table=11
/product="6-Phosphogluconate Dehydrogenase"
/protein_id="AAC67654.1"
/db_xref="GI:3328457"
/translation="WAPNTDIGLIGLAVAGKNVLNMVHDGFSVSVYNSRPAKTEEPL
KDHSGLAOGFTTIOEFVQSLKRPRIKIMIKGAPVDEMIALPLPELGGDILIDG
GNSYLYDSEORVYDLKKEGILFVGVSGEGARKGPSIMPGNIDMAPIAPIFOS
IAAOYDGRCCSMWIGTGAGHGVKAVHNGITGQICETIYELIKTRNLSTEOGN
IPEFNQTDNLSTYLGAANAVALIADENGNAIASIILIDACOKGCRVAVDAIKAGV
PMSLTIESVLARYLSTKKEVTRKAQEPFGIPLDQPPQEAATIEDVREALYAKIIL
SYAOGFMLKQVSDQKWDNLNGELALIRGCGCIIOAFLDKIHOGFENSPHSLIL
QDYEFKVLDESETGFRRAVLAHIGSVAIPCLSSALSFYDGYRTVDSLSLTVQGLRDY
FGAHGERRDDCPGEGFYHTDMLTETKTRFY"
complement(3757. .5565)
/gene="lepA"
/note="CT064"
complement(3757. .5565)
/gene="lepA"
/codon_start=1
/transl_table=11
/product="GTPase"
/protein_id="AAC67655.1"
/db_xref="GI:3328458"
/translation="MKPKIENIRNFSTIAIHDKGSTIADRLLESTSTIEQREMQ
LLDSMDLEREGITIKAHVPMTYEYEGEYELNIDTPGHVDFSEYVSRSLAACEGA
LLIYDAAQVOAOSLANVYLALERLDFTIPLVNLKIDLPAAQPEAIKROIEEFIGLDT
NTIACSAKTGGCIPRILSIIIRLVPPRPPOETELKALIPDSHPYIGIYVYVVIS
GEIKGDRITPMATKGSSEVYIGIAFLPEATLMGSLRACOVGFIANLKKVVKIKI
GDVYTVVHPAKPEPLEGFEKRIKPVVFAGITPDDSDFTLKDALGRLQNLNSALIEQ
ENSHSLGFGFCGFLGLHELIIEFIRISREFDLDIATAPSYKVVLKNSKTLFIDN
PTAYADPALIEHMEEPYHVNITPOEYLSNMSLCMDKRGICLCTMDLMDLRILASY
ELPLNEIYSDNDEKLKSVTKGYSDFYDLGDKKCAIILKELLINDEAVDFSCVYHR
DKAESKRSICEKLVDTIPROLEKTIPIQMAINKKIIAETIRALAKNVTACYGGDIT
RKKRLMDQKKGKKRMKEFGKVSIPNTAFVEVLMKE"
complement(5751. .7337)
/gene="CT065"
complement(5751. .7337)
/codon_start=1
/transl_table=11
/product="ADP/ATP Translocase"
/protein_id="AAC67656.1"
/db_xref="GI:3328459"
/translation="MTQFAKKPEGLRSEFLMPIHMHLEKVLPMPLMFCSTSPNTTL
RDTKDTLVTAPGSAEAIPIPKIMLVYPSAVVFLITAKLSNLIKALFPFVAVISPF
VVFRLPEVVIYPCRHILHPTAFADTDSILPSCGMGFIAMLRNMTFAVFYVLSLWG
SVSLIMFWGFANETIKISEAKRFALGVGAVNALLISGPAIIMSSKLRASLGGCAG
PWGVSYFLMAMFLCSCAIIACACYMMNRYVLTDRFYNPAELKAKSKPKMSKESEF
SYLRSPYMLLALLVICYICINIVEYTWKSQULKMGFPNPNDYSAFNGNFSFWTVYV

gene 8082. .9802
/gene="ytdA"
/note="CT067"
8822. .9802
/gene="ytdA"
/codon_start=1
/transl_table=11
/product="Solute Protein Binding Family"
/protein_id="AAC67658.1"
/db_xref="GI:3328461"
/translation="WSEFTRKRYKILRLGLCLACGLFMSNCSRSSRGNOQADESTIYL
SNMRICDCVSRITGDYRKNTIVLIDGALDIPSYSEVWKDEDEBMAASQILFNGGLEH
SASLRKHLGKNPVYDLGQRLNKNCFELSLSEEGPPDPHTIDMKVMVMAAKVMAAL
IQOFOYEDDFQKANDQILSMEELDRMAASLSGTPENKRYLVYGHNAFSYFTRRYL
SSDERVSGENSRISIPESGISPEAOISIRIMRIVEYISANDVEYVLEPTLQNDAL
RKVYCSGSGKIRLAKRPLKSDVNCNDFSTFQHNVTITBELGLTYLE"
9795. .10574
/gene="ytdB.1"
/note="CT068"
9795. .10574
/gene="ytdB.1"
/codon_start=1
/transl_table=11
/product="rRNA methylase"
/protein_id="AAC67659.1"
/db_xref="GI:3328462"
/translation="MNRDMAILMSVEDLCYVNDSDVLCITFTESLPAGMAAIIIGPG
AGSKTLKASGLIRASSGOSLIFGQRSKYHHRATVYPOASVDMDPMPVLDVLM
GCTGKGIWNRISTDDROEMKRLIERVGLFANQIGKLSGGQOQRAFLARSLMOKA
DVLMDLEFSAIDMAVQMVVDVQELKSEKTIYVHHDSLNVKRLFDHVLINKHL
VCSGVEECLTKEALFQAYGCELELDYTLKLSRKYOGSC"
10575. .11930
/gene="ytdC"
/note="CT069"
10575. .11930
/gene="ytdC"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAC67660.1"
/db_xref="GI:3328463"
/translation="MLSCIFQDTTIFLSSFLAVSLICMTALTALMGTTLLVEROPLSESL
SHACYPGLLIGALLSYKVPARFDSLMVNIIFPGCLASVAGCGISFLEKELAMHDSAL
CLIVAFPGVAVIIVSYKDCCLPLYNKINYLGOAATLIGYTEKALAILFCSAVY
LMMYROIISVAIPREFAVSCGLRTRTELIVLVIFISLIVYSGRSVGLIILISAPVA
PPLSARQSDRLSTLILSSIFGICGALGCFPSVAFPCQIVVEGKRPSTIILPGPLY
VFRGAVLFLCIFSWKTGMITRYFRKRWPLFSRDEHLKLFMWLRQONTVOGMRD
FVRSKRQYQYEGDKVFPFRPRMELCKGLVCSBHOWSLTPKGLARAARAKIVARHME
SYLVSQLDFNNEVHHFAEMEHLVTDLDSTLSQMLQDDPYDHPHOEIRPKTRKSDG
C"
11920. .12876
/gene="ytdD"
/note="CT070"
11920. .12876
/gene="ytdD"
/codon_start=1
/transl_table=11
/translation="MTQFAKKPEGLRSEFLMPIHMHLEKVLPMPLMFCSTSPNTTL
RDTKDTLVTAPGSAEAIPIPKIMLVYPSAVVFLITAKLSNLIKALFPFVAVISPF
VVFRLPEVVIYPCRHILHPTAFADTDSILPSCGMGFIAMLRNMTFAVFYVLSLWG
SVSLIMFWGFANETIKISEAKRFALGVGAVNALLISGPAIIMSSKLRASLGGCAG
PWGVSYFLMAMFLCSCAIIACACYMMNRYVLTDRFYNPAELKAKSKPKMSKESEF
SYLRSPYMLLALLVICYICINIVEYTWKSQULKMGFPNPNDYSAFNGNFSFWTVYV

```

/product="Integral Membrane Protein"
/db_xref="GI:3328464"

Query Match          2.4%; Score 39; DB 1; Length 14168;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 AAGAAAGTTCTGCCAATGTTCTTAATGTTCTGTATTT 161
      |||||||
Db 7265 AAGAAAGTTCTGCCAATGTTCTTAATGTTCTGTATTT 7227

RESULT 7
AX281684 5032 bp DNA linear PAT 02-NOV-2001
LOCUS AX281684
DEFINITION Sequence 93 from Patent WO0177389.
ACCESSION AX281684
VERSION AX281684.1 GI:16608935
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 Shiffman,D., Somogyi,R., Lawn,R., Sellhammer,J.J., Porter,G.J.,
TITLE Mikita,T. and Tai,J.
JOURNAL Genes expressed in foam cell differentiation
Patent: WO 0177389-A 93 18-Oct-2001;
FEATURES Incyte Genomics, Inc. (US)
SOURCE Location/Qualifiers
1. 5032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 445186.7"

BASE COUNT 1473 a 1026 c 1055 g 1477 t 1 others
ORIGIN

Query Match          1.3%; Score 21; DB 6; Length 5032;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 700 TAGATCCTTGGGGAATTTCTT 720
      |||||||
Db 2927 TAGATCCTTGGGGAATTTCTT 2947

RESULT 8
AR064709 8220 bp DNA linear PAT 29-SEP-1999
LOCUS AR064709
DEFINITION Sequence 11 from patent US 5849306.
ACCESSION AR064709
VERSION AR064709.1 GI:5994925
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 8220)
TITLE Sim,K.Lee., Chichnis,C., Miller,L.H., Peterson,D.S., Su,X.-Z. and
Willems,T.E.
JOURNAL Binding domains from Plasmodium vivax and Plasmodium falciparum
erythrocyte binding proteins
Patent: US 5849306-A 11 15-DEC-1998;
FEATURES Location/Qualifiers
1. 8220
/organism="unknown"

BASE COUNT 3194 a 1249 c 1777 g 2000 t
ORIGIN

Query Match          1.3%; Score 21; DB 6; Length 8220;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 64 AAGAAACCTTTTGGAAT 84
      |||||||
Db 5046 AAGAAACCTTTTGGAAT 5066

RESULT 9
AR089156 8220 bp DNA linear PAT 07-SEP-2000
LOCUS AR089156
DEFINITION Sequence 11 from patent US 5993827.
ACCESSION AR089156
VERSION AR089156.1 GI:10015913
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 8220)
TITLE Sim,K.Lee., Chichnis,C., Miller,L.H., Peterson,D.S., Su,X.-Z. and
Willems,T.E.
JOURNAL Binding domains from Plasmodium vivax and Plasmodium falciparum
erythrocyte binding proteins
Patent: US 5993827-A 11 30-NOV-1999;
FEATURES Location/Qualifiers
1. 8220
/organism="unknown"

BASE COUNT 3194 a 1249 c 1777 g 2000 t
ORIGIN

Query Match          1.3%; Score 21; DB 6; Length 8220;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 AAGAAACCTTTTGGAAT 84
      |||||||
Db 5046 AAGAAACCTTTTGGAAT 5066

RESULT 10
AL592487 17462 bp DNA linear PRI 15-NOV-2001
LOCUS AL592487/c
DEFINITION Human DNA sequence from clone Rpl1-290b1 on chromosome 1, complete
sequence.
ACCESSION AL592487
VERSION AL592487.10 GI:16973138
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 17462)
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:16416250.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

```

was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chrl>
Rpl1-290B1 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: PBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-290B1 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone Rpl1-290B1 is at 1 in this sequence. The true left end of clone Rpl1-474021 is at 15463 in this sequence.

FEATURES

source
1..17462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="Rpl1-290B1"
/clone_lib="RPCI-11.2"
BASE COUNT 4402 a 4424 c 4149 g 4487 t
ORIGIN

Query Match 1.3%; Score 21; DB 9; Length 17462;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 353 CTTTAAATTTCTTGCCCT 373
|||||
Db 15493 CTTTAAATTTCTTGCCCT 15473

RESULT 11
PEAVARIA
LOCUS 19124 bp DNA linear INV 14-SEP-1995
DEFINITION Plasmodium falciiparum (strain Dd2) variant-specific surface protein (var-1) gene, complete cds.
ACCESSION LA0608
VERSION 1
KEYWORDS variant-specific surface protein.
SOURCE Plasmodium falciiparum (strain Dd2) blood stage DNA.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 19124)
Su,X.Z., Heatwole,V.M., Wertheimer,S.P., Guinet,F., Herfeldt,J.A., Peterson,D.S., Ravetch,J.A. and Wellems,T.E.
The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciiparum-infected erythrocytes
JOURNAL Cell 82 (1), 89-100 (1995)
MEDLINE 95330813
PUBMED 7606788
FEATURES

source
Location/Qualifiers
1..19124
/organism="Plasmodium falciiparum"
/strain="Dd2"
/db_xref="taxon:5833"
/def_stage="blood stage"
2068..2967
/note="putative"
/label="rif_homology
7307..17609
/gene="var-1"
Join((7307..15139,16206..17609)
/gene="var-1"
/note="blood stage; PfEMP1 gene related; putative"
/codon_start=1
/product="variant-specific surface protein"
/protein_id="AA7536.1"
/db_xref="GI:886375"
/translation="MVELAKMGPEKAAGDDIEDSAKHMFDRIKGVYDKVEEAK
RGKIGRLSEAKFEKNESDPTEDCDLHKYHTVNTVINPCADRSDFRSDEY
GGCCTHNRIKDSOGDNKACAPYRRLHVCDDNLEOIEPIKITNTNHLNLDVCAAKF

EGSITDQPYKATVGDSPSICITMLARSPADIGDLYVRGDLVLGNPOETKOROLE
MNTKTFGKIYEKINQAEARGNDQPEPPKILBEDMTNABETVMKATLTCNMGNYPFA
TCNGERTKGYCRNDQOVPTPEYVQYILMEFEAADEFSRKNKATIKVGRNCRK
DKEDKDRCSNGIDCEKTKRAIGLRKGCISCLACNPTVYM INNOEBOPOKOK
KYEIEIKKYENGASGSRKADAGTTTNTNDEYKFEYBELANSERYDKEFLKLS
NEEICTIVKDEGGTIDEKVNNSDSTGASTANESGCTEYRSKYCPCGVKYN
NGSSNMEERKNKCSKGLYEKPKEGTITILSGKHDDIEERLCKCEKNG
DTINGSGSGTGGSGNSGROLEYEKKYKGVENVYGHDEDEEYEVANKNAGLIC
ILKNOKKNEGGGNTSEKDEDEIKTNPFTTYVAMHLKDSIMKKKLORCLONGNR
IKCGNNKCNDCBECFKRMITOKRDEWKIVHFPTQNTKGGSDNTAEILPEHDVY
LOYNLOEFLKGDSEDAESEKSNLSAEAEELKHLREIIESDNNNOEASVGGVE
OKNIMDKLLANEKDEADLCLEIHEDEEEKKCGNGCIEEGENRPNPSCGESNRY
YPLANKVAYOMHHKAKTOLASRAGRLARDITLAFONRNSGTJLKGOICKINENY
SNDGRNSGGTCTGDDHGGYRMRITGEGKGTSTTKKNVFLPBRPHMTSL
ENLDVGYTKNDKASHSLDGVLAATDMAEIIKKRKDONNIDLTPTIIOKODEAK
RAVRYSFADLDLIRGDMWDEKSSYDMEIRLITVRNKIEKHGDIKDNKTYGDS
KAPVYKRLRADWMEANRQVWRAMKCATKIGICGMVDDVITPOLRMNEMAWMYK
AOSOEYDKLKKICADCSKSGDKCTGQDVDCGKCAACDKYKEIEKEMQMRISDK
YNLYIAKTTSTNPRGVLGDDDPDYVOOMVFLPTIHKASIAARVLYKRAASPTET
AAAPITPYSSTAAGYTHOETCYGCCOCTOPECKKHGATSTSTTKENKTYRQPPPE
YATACDCINRSQTEEPKKKEENVESACKIYKILLEGKNGRTTVECPNKESYPIWDCK
NNTIDISHGACMPPRKOKLCLYVIAHSEOTENITDNDLDAEITKTAAEFTLSYTY
KSKNDSEAKILDRGLIPSOFLRSMVYTFGDRDILNTDLSKKNQDAKADKIGKFE
SKDSKSPSGLSROEMKTNPEIWMGMLCATKYVDTDRKRLKNDYSDKYNOQO
NGNPSLEEFKAPQFLRMILMEGEFCAEROKKENITKDACNEINQOCCDAKHRCN
OACRAYOYEVENKREKFEFGOTNNFVLKANOPODEYKGYEKGYKQPIQNEVLLQK
CDNNKCSMDGNVLSVSEKPEFKYAHKYPEKDCCTOGKHPSTIPPEPVYQPEA
PTVTVDCSIVKTLFKDPTNNSDACGLKYGTAPSVMKCIPTSDRSGAGATGKSGSD
SGSICIPPRRRRLVYGLQEWATLAPGEGAPSHSRADLRNAPFIOAALETFFLMD
RYKEEKKPOGDSOQALSOULTSYSDDEEDPDRLONGKLPDFLRIMEFTLDGYRD
ILVHGNTSDSGNTGNSNNNIIVLEASGNKEDMOKIOEKIEOILPKNGGPTLVSSA
OTPDKMNNEHAEISTKMGICALTYTERKNPTKNSPDKASDENKIEKDEVEYKPEFGSADH
GTASTPTGTYTQYDIEKVKLEDTSGKATPSASDPTLDSFVLPPTFRILEMGON
FCRRKRHLQIHECVVEENGGSRRGCTIROYSGOEAENMLPKRNDGVPLDEXP
SCAPCSYRKMIESKGEVEKOEKRAYEOQDKCVNOSNKNHDFCETLITTSVAKNF
LKTGPKCPNVNVEKTIPTDDKRTKHKRDCPCCLFAGNCFEGCIRDEMDGCRNKS
IDATDIEGNGVSTVLEMRKVSADSKSGFNGDLEMACGAGLFEIGIRDEMDGCRNKS
VYCKPEMNGKAGKHIIQIRALYKRWYVEFEFYKIKKHISRINGETSPICIKNC
VEKWVDOKRKRMEKITERKQOYKNDSDDDNVSFETLIPOTTDANKAKYIKLSK
FGNSGCSASANEONKGETKDAIDCMKRLKDKIGEEKHNHGTSDIESCTQOPOT
LEDFTLDDIDLETEBAKNNMPKICEVNLKTAQODEGCVPAENSEEPAVDSKERTP
EOTPVLLKPEEAVAPPPPEPPEKAPAPIPOPPPTPTOTLLDNPVHLYALVSTLAW
SVGIGFAFTFYFLYKTKSSVGNLPIPOKSDYDIPKLSNRYIPIYSGKYGK
RYTILEGSGTDSGYTDHYSIDTSSSEYEEMDINDIYVSPSPKTLIEVLYEGRK
NFTVASGNMTASGNWTASGNTPSPDONDONCIGISSKTTDEMNOJADEPISQTL
OSEPNTERNMLGTVNDNTHPTTSHHVEEKFPTMSIHDRLSEGEENYNDMFGSNL
PINISDSTNSMDSLTSNNHSPYNDKNDLYSIDLINDALSGNHLIDYDEMILPKREL
FGTHNHTKHTNTYVAVAKPARDDPTTNQINLPHKMLDRHMCERKMNHHERLPKLEL
WENETHSGDINSIGIPSGNHVLANPTVNSIOIDMDNKTKNEITNMPTNDPKSTMDTILDD
LEKYNPEYVDFEYDDIILYHDVDEKSMMDIYVDHNNTNNMNDVPTKMHIEKNITV
NKKETFEFEYPISDIWNIT"

BASE COUNT 7824 a 2190 c 2790 g 6320 t
ORIGIN

Query Match 1.3%; Score 21; DB 3; Length 19124;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 64 AAGAAAACCTTTGGAAT 84
|||||
Db 12366 AAGAAAACCTTTGGAAT 12386

RESULT 12
AR089157


```

* 23207 23638: contig of 432 bp in length
*      gap of unknown length
* 23639 24265: contig of 627 bp in length
*      gap of unknown length
* 24266 25200: contig of 935 bp in length
*      gap of unknown length
* 25201 26153: contig of 953 bp in length
*      gap of unknown length
* 26154 26999: contig of 846 bp in length
*      gap of unknown length
* 27000 27419: contig of 420 bp in length
*      gap of unknown length
* 27420 28185: contig of 766 bp in length
*      gap of unknown length
* 28186 29128: contig of 943 bp in length
*      gap of unknown length
* 29129 29817: contig of 689 bp in length
*      gap of unknown length
* 29818 30211: contig of 394 bp in length
*      gap of unknown length
* 30212 30820: contig of 609 bp in length
*      gap of unknown length
* 30821 31969: contig of 1149 bp in length
*      gap of unknown length
* 31970 32714: contig of 745 bp in length
*      gap of unknown length
* 32715 34020: contig of 1306 bp in length
*      gap of unknown length
* 34021 35177: contig of 1157 bp in length
*      gap of unknown length
* 35178 36602: contig of 1425 bp in length
*      gap of unknown length
* 36603 37916: contig of 1314 bp in length
*      gap of unknown length
* 37917 38222: contig of 306 bp in length
*      gap of unknown length
* 38223 39014: contig of 792 bp in length
*      gap of unknown length
* 39015 40181: contig of 1167 bp in length
*      gap of unknown length
* 40182 41359: contig of 1178 bp in length
*      gap of unknown length
* 41360 42186: contig of 827 bp in length
*      gap of unknown length
* 42187 42987: contig of 801 bp in length
*      gap of unknown length
* 42988 44029: contig of 1042 bp in length
*      gap of unknown length
* 44030 45441: contig of 1412 bp in length
*      gap of unknown length
* 45442 46864: contig of 1423 bp in length
*      gap of unknown length
* 46865 48255: contig of 1391 bp in length
*      gap of unknown length
* 48256 49365: contig of 1110 bp in length
*      gap of unknown length
* 49366 50340: contig of 975 bp in length
*      gap of unknown length
* 50341 51309: contig of 969 bp in length
*      gap of unknown length
* 51310 52722: contig of 1413 bp in length
*      gap of unknown length
* 52723 54001: contig of 1279 bp in length
*      gap of unknown length
* 54002 55737: contig of 1736 bp in length
*      gap of unknown length
* 55738 57052: contig of 1315 bp in length
*      gap of unknown length
* 57053 58809: contig of 1757 bp in length
*      gap of unknown length
* 58810 60365: contig of 1556 bp in length
*      gap of unknown length
* 60366 62096: contig of 1731 bp in length

```

```

* 62097 64348: gap of unknown length
*      contig of 2252 bp in length
* 64349 65921: gap of unknown length
*      contig of 1573 bp in length
* 65922 67231: gap of unknown length
*      contig of 1310 bp in length
* 67232 69525: gap of unknown length
*      contig of 2294 bp in length
* 69526 70784: gap of unknown length
*      contig of 1259 bp in length
* 70785 73256: gap of unknown length
*      contig of 2472 bp in length

Query Match      1.3%  Score 21; DB 2; Length 90106;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATACAAAGACACTTATTG 207
Db 45690 ATACAAAGACACTTATTG 45670
|||||
RESULT 14
AC092392 98286 bp DNA linear PRI 15-DEC-2001
LOCUS Homo sapiens chromosome 1 clone RP4-544H6, complete sequence.
AC092392 AL109625
DEFINITION AC092392.2 GI:17861066
VERSION AC092392.2 GI:17861066
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 98286)
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
Haugen, E.D.
TITLE Direct Submissio
JOURNAL Submitted (03-JUL-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 98286)
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
Haugen, E.D.
TITLE Direct Submissio
JOURNAL Submitted (15-DEC-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Dec 15, 2001 this sequence version replaced gi:14589610.
COMMENT ----- Genome Center
Center: University of Washington Genome Center
Center Code: UMGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP4-544H6 (sc0188)
----- Summary Statistics
Sequencing vector: plasmid; 22% of reads
Sequencing vector: plasmid; 78% of reads
Chemistry: Dye-terminator ET; 32% of reads
Chemistry: Dye-terminator Big Dye; 68% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 98278 bases at least Q40
Consensus quality: 98286 bases at least Q30
Insert size: 98287; sum-of-ctrls
Quality coverage: 9.1x in Q20 bases; sum-of-ctrls
-----
Overlapping Sequences:

```

5': RP5-1043L3 (UMGC:sc0545) AC099062
3': RP4-671122 (UMGC:sc0205) AL356298

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector. In order to accurately represent the entire circular BAC, small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII

EcoRI

HindIII

SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint
-----	-----	-----	-----	-----	-----
6023	5830	8696	8978	1849	1934
-----	-----	-----	-----	-----	-----
2067	2146	6101	6394	6382	6210
-----	-----	-----	-----	-----	-----
7084	7102	6972	6986	512	<800
-----	-----	-----	-----	-----	-----
2144	2146	6986	6986	449	<800
-----	-----	-----	-----	-----	-----
1462	1431	11036	10738	1792	1934
-----	-----	-----	-----	-----	-----
2877	2772	826	827	7699	7751
-----	-----	-----	-----	-----	-----
252	<800	1438	1380	3248	3382
-----	-----	-----	-----	-----	-----
547	<800	2212	2194	6225	6210
-----	-----	-----	-----	-----	-----
815	883	1590	1521	9	<800
-----	-----	-----	-----	-----	-----
308	<800	3225	3154	720	726
-----	-----	-----	-----	-----	-----
515	<800	3261	3312	1705	1735
-----	-----	-----	-----	-----	-----
1321	1320	671	<800	2231	2162
-----	-----	-----	-----	-----	-----
11034	10897	2171	2194	83	<800
-----	-----	-----	-----	-----	-----
4803	4847	3889	3911	298	<800
-----	-----	-----	-----	-----	-----
1643	1622	2401	2382	307	<800
-----	-----	-----	-----	-----	-----
2140	2146	742	827	3417	3382
-----	-----	-----	-----	-----	-----
2925	2947	8122	8089	2700	2801
-----	-----	-----	-----	-----	-----
4658	4573	988	974	262	<800

-----	-----	-----	-----	-----	-----
3569	3631	2263	2194	1407	1374
-----	-----	-----	-----	-----	-----
3936	4009	825	827	5580	5488
-----	-----	-----	-----	-----	-----
653	<800	7300	7287	2031	2162
-----	-----	-----	-----	-----	-----
18	<800	1900	1898	2574	2646
-----	-----	-----	-----	-----	-----
1053	1111	4808	4762	2085	2162
-----	-----	-----	-----	-----	-----
1135	1111	2669	2653	897	907
-----	-----	-----	-----	-----	-----
50	<800	2712	2746	202	<800
-----	-----	-----	-----	-----	-----
113	<800	6452	6394	706	<800
-----	-----	-----	-----	-----	-----
2617	2615	2160	2194	3060	3207
-----	-----	-----	-----	-----	-----
689	<800	3777	3706	3662	3618
-----	-----	-----	-----	-----	-----
402	<800	792	<800	1702	1655
-----	-----	-----	-----	-----	-----
7554	7725	-----	-----	250	<800
-----	-----	-----	-----	-----	-----
592	<800	-----	-----	1010	1005
-----	-----	-----	-----	-----	-----
5487	5529	-----	-----	333	<800
-----	-----	-----	-----	-----	-----
1643	1622	-----	-----	233	<800
-----	-----	-----	-----	-----	-----
6052	6239	-----	-----	1223	1200
-----	-----	-----	-----	-----	-----
896	883	-----	-----	4551	4630
-----	-----	-----	-----	-----	-----
5326	5322	-----	-----	1948	1934
-----	-----	-----	-----	-----	-----
1036	1030	-----	-----	4120	3964
-----	-----	-----	-----	-----	-----
6826	6811	-----	-----	7902	8074
-----	-----	-----	-----	-----	-----
1378	1320	-----	-----	5161	5086
-----	-----	-----	-----	-----	-----
599	<800	-----	-----	543	<800
-----	-----	-----	-----	-----	-----
2743	2615	-----	-----	6163	6210
-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	5260	5086
-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	1692	1655
-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	2232	2162
-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	451	<800
-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	119	<800

FEATURES
source Location/Qualifiers

1. 98286
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RPCI human PAC library 4"
ORIGIN
BASE COUNT 27438 a 19511 c 20578 g 30759 t

Query Match 1.3%; Score 21; DB 9; Length 98286;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 TAGATCCTTGGGAGATTCTT 720
|||||

Db 60168 TAGATCCTTGCGGATTCTT 60188

RESULT 15

AC113803

LOCUS

AC113803 126996 bp DNA linear HTG 17-JUL-2002

DEFINITION Rattus norvegicus clone CH230-293F17, *** SEQUENCING IN PROGRESS

***, 52 unordered pieces.

ACCESSION

AC113803 GI:21744586

VERSION

HTG: HTGS_PHASE1.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 126996)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Boyle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

Burch,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,

Homi,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlssoon,E., Kelly,S., Khan,U., King,J., Korvah,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Louised,H.,

Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,N., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokkwo,S., Ogun,M., Okunnu,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

Rivers,M., Rojas,A., RojuboKan,I., Rolfe,M., Ruiz,S., Savery,G.,

Scherrer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,

Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,

Umsani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,

Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,

Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Weinstock,G. and Gibbs,R.

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GSHO
 Center clone name: CH230-293F17
 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 72236 bases at least Q40
 Consensus quality: 75069 bases at least Q30
 Consensus quality: 77416 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 52 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1036: contig of 1036 bp in length
 * 1037 1136: gap of unknown length
 * 1137 2213: contig of 1077 bp in length
 * 2214 2313: gap of unknown length
 * 2314 3997: contig of 1684 bp in length
 * 3998 4097: gap of unknown length
 * 4098 5272: contig of 1175 bp in length
 * 5273 5372: gap of unknown length
 * 5373 6575: contig of 1203 bp in length
 * 6576 6675: gap of unknown length
 * 6676 8157: contig of 1481 bp in length
 * 8157 8257: gap of unknown length
 * 8257 9450: contig of 1194 bp in length
 * 9451 9551: gap of unknown length
 * 9551 10960: contig of 1410 bp in length
 * 10961 11060: gap of unknown length
 * 11061 12118: contig of 1058 bp in length
 * 12119 12218: gap of unknown length
 * 12219 13878: contig of 1660 bp in length
 * 13879 13978: gap of unknown length
 * 13979 15252: contig of 1274 bp in length
 * 15253 15352: gap of unknown length
 * 15353 16951: contig of 1599 bp in length
 * 16952 17051: gap of unknown length
 * 17052 18138: contig of 1087 bp in length
 * 18139 18238: gap of unknown length
 * 18239 19414: contig of 1176 bp in length
 * 19415 19514: gap of unknown length
 * 19515 21376: contig of 1862 bp in length
 * 21377 21476: gap of unknown length
 * 21477 23039: contig of 1553 bp in length
 * 23040 23139: gap of unknown length
 * 23140 25010: contig of 1871 bp in length
 * 25011 25110: gap of unknown length
 * 25111 26280: contig of 1170 bp in length
 * 26281 26380: gap of unknown length
 * 26381 27993: contig of 1613 bp in length
 * 27994 28093: gap of unknown length
 * 28094 30038: contig of 1945 bp in length
 * 30039 30138: gap of unknown length
 * 30139 31754: contig of 1616 bp in length
 * 31755 31854: gap of unknown length
 * 31855 33816: contig of 1962 bp in length
 * 33817 33916: gap of unknown length
 * 33917 36417: contig of 2501 bp in length
 * 36418 38336: gap of unknown length
 * 38337 38436: contig of 1819 bp in length
 * 38436: gap of unknown length

COMMENT

On Jul 13, 2002 this sequence version replaced gi:19525967.

Genome Center

```
* 38437 39646: contig of 1210 bp in length
* 39647 39746: gap of unknown length
* 39747 41954: contig of 2208 bp in length
* 41955 42054: gap of unknown length
* 42055 44188: contig of 2134 bp in length
* 44189 44288: gap of unknown length
* 44289 47196: contig of 2908 bp in length
* 47197 47296: gap of unknown length
* 47297 49882: contig of 2586 bp in length
* 49883 49982: gap of unknown length
* 49983 51747: contig of 1765 bp in length
* 51748 51847: gap of unknown length
* 51848 54070: contig of 2223 bp in length
* 54071 54170: gap of unknown length
* 54171 55692: contig of 1522 bp in length
* 55693 55792: gap of unknown length
* 55793 58784: contig of 2992 bp in length
* 58785 58884: gap of unknown length
* 58885 60823: contig of 1939 bp in length
* 60824 60923: gap of unknown length
* 60924 62885: contig of 1962 bp in length
* 62886 62985: gap of unknown length
* 62986 64984: contig of 1999 bp in length
* 64985 65084: gap of unknown length
* 65085 67271: contig of 2187 bp in length
* 67272 67371: gap of unknown length
* 67372 70966: contig of 3595 bp in length
* 70967 71066: gap of unknown length
* 71067 74043: contig of 2977 bp in length
* 74044 74143: gap of unknown length
* 74144 77759: contig of 3616 bp in length
* 77760 77859: gap of unknown length
* 77860 79868: contig of 2009 bp in length
* 79869 79968: gap of unknown length
* 79969 83800: contig of 3832 bp in length
* 83801 83900: gap of unknown length
* 83901 86500: contig of 2600 bp in length
* 86501 86600: gap of unknown length
* 86601 89461: contig of 2861 bp in length
* 89462 89561: gap of unknown length
* 89562 93647: contig of 4086 bp in length
* 93648 93747: gap of unknown length
* 93748 97544: contig of 3797 bp in length
* 97545 97644: gap of unknown length
* 97645 101973: contig of 4329 bp in length
* 101974 102073: gap of unknown length
* 102074 106111: contig of 4038 bp in length
* 106112 106211: gap of unknown length
* 106212 112152: contig of 5941 bp in length
* 112153 112252: gap of unknown length
* 112253 116920: contig of 4668 bp in length
* 116921 117020: gap of unknown length
* 117021 120957: contig of 3937 bp in length
* 120958 121057: gap of unknown length
* 121058 126996: contig of 5939 bp in length.
```

```
FEATURES
    source
        1. .126996
        /organism="Rattus norvegicus"
        /db_xref="taxon:10116"
```

```
Query Match 1.3%, Score 21: DB 2: Length 126996;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 1576 ATTCAGCTCTGCTCTTCAT 1596

Db 43370 ATTCAGCTCTGCTCTTCAT 43390

```
RESULT 16
AC007335/c AC007335 134482 bp DNA linear PRI 03-NOV-2000
LOCUS Homo sapiens chromosome 16 clone RP11-165M2, complete sequence.
DEFINITION
```

```
ACCESSION AC007335
VERSION AC007335.6 GI:11079401
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
```

```
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 134482)
JOURNAL DOE Joint Genome Institute.
AUTHORS Sequencing of Human Chromosome 16
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 134482)
Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meircke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
```

```
TITLE Direct Submission
JOURNAL Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
REFERENCE Genome Institute, Los Alamos National Laboratory, MS M888, Los
AUTHORS Alamos, NM 87545, USA
TITLE 3 (bases 1 to 134482)
REFERENCE DOE Joint Genome Institute.
```

```
JOURNAL Direct Submission
AUTHORS Submitted (03-NOV-2000) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 3, 2000 this sequence version replaced gi:9795550.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
```

```
FEATURES
    source
        1. .134482
        /organism="Homo sapiens"
        /db_xref="taxon:3606"
        /chromosome="16"
        /clone="RP11-165M2"
```

```
BASE COUNT 35905 a 28147 c 29995 g 40435 t
ORIGIN
```

```
Query Match 1.3%, Score 21: DB 9: Length 134482;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 187 ATCAAAAGACACTTATTG 207

Db 78708 ATCAAAAGACACTTATTG 78688

```
RESULT 17
AC104838 143279 bp DNA linear PRI 31-MAY-2002
LOCUS Homo sapiens chromosome 1 clone RP4-671I22, complete sequence.
ACCESSION AC104838 AL356298
VERSION AC104838.2 GI:21281548
KEYWORDS HTG.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 143279)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
and Haugen,E.D.
```

```
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143279)
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Center, University of Washington,
REFERENCE Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 143279)
```

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saepthimachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C., and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAY-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 COMMENT On May 31, 2002 this sequence version replaced gl:17976476.
 ----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UMG
 Web site: <http://www.genome.washington.edu>
 Contact: uwgchits@u.washington.edu
 Drafting Center: SC
 ----- Project Information
 Center project name: chr-1
 Center clone name: RP4-671122 (sc0205)
 ----- Summary Statistics
 Sequencing vector: plasmid; 70% of reads
 Sequencing vector: plasmid; 108752; 30% of reads
 Chemistry: Dye-terminator ET; 59% of reads
 Chemistry: Dye-terminator Big Dye; 41% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 143182 bases at least Q40
 Consensus quality: 143276 bases at least Q30
 Consensus quality: 143278 bases at least Q20
 Insert size: 143279; sum-of-contigs
 Quality coverage: 9.3x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:
 5': Mapping in progress
 3': RP11-442N6 (UMC:sc0307) AC103552, 43012-bp overlap

----- Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

----- EcoRI

----- HindIII

----- BglII

----- SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

----- 9950 10064 1637 1634 8109 8153
 ----- 2184 2191 449 <800 5671 5728
 ----- 8065 8142 512 <800 8094 8153

----- 486 <800 2814 2847 592 <800
 ----- 106 <800 1247 1206 5487 5342
 ----- 7113 7150 4695 4723 1643 1649
 ----- 988 992 3373 3403 7390 7505
 ----- 2263 2191 6390 6351 896 888
 ----- 825 842 1948 1923 5327 5342
 ----- 8638 8681 4120 1036 1094 1094
 ----- 1900 1918 9240 9293 6826 6849
 ----- 4809 4785 5162 5133 1378 1354
 ----- 2669 2648 543 <800 599 <800
 ----- 2712 2773 6163 6351 2743 2754
 ----- 6452 6486 5260 5133 3995 3978
 ----- 2160 2191 1692 1634 1146 1094
 ----- 3777 3727 2232 2178 7 <800
 ----- 4526 4454 451 <800 1613 1649
 ----- 4195 4163 119 <800 3589 3609
 ----- 117 <800 17870 17825 6332 6421
 ----- 273 <800 440 <800 6781 6849
 ----- 3560 3601 844 843 215 <800
 ----- 5446 5428 670 <800 6 <800
 ----- 6564 6486 655 <800 1740 1649
 ----- 697 <800 849 843 3656 3609
 ----- 10283 10064 4779 4723 1657 1649
 ----- 720 <800 343 <800 1875 1868
 ----- 4163 4163 779 763 4873 4794
 ----- 1386 1434 6371 6351 826 823
 ----- 1472 1434 1554 1513 720 728
 ----- 1214 1182 221 <800 321 <800
 ----- 678 <800 4724 4723 69 <800
 ----- 50 <800 1220 1206 1820 1785
 ----- 3978 3963 3375 3403 3084 3148
 ----- 6193 6203 384 <800 1525 1492
 ----- 76 <800 5108 5133 4613 4568
 ----- 2441 2474 686 <800 818 823
 ----- 1446 1434 764 763 3366 3409
 ----- 2575 2648 13713 13478 7051 7129
 ----- 4884 4914 3922 4010 2376 2386


```

* 111748 111847: gap of unknown length
* 111848 115180: contig of 3333 bp in length
* 111849 115280: gap of unknown length
* 115281 118548: contig of 3268 bp in length
* 118549 118648: gap of unknown length
* 118649 121539: contig of 2891 bp in length
* 121540 121639: gap of unknown length
* 121640 124526: contig of 2687 bp in length
* 124527 124627: gap of unknown length
* 124627 127512: contig of 2886 bp in length
* 127513 127613: gap of unknown length
* 127613 131027: contig of 3415 bp in length
* 131028 131127: gap of unknown length
* 131128 133604: contig of 2477 bp in length
* 133605 133704: gap of unknown length
* 133705 135894: contig of 2190 bp in length
* 135895 135994: gap of unknown length
* 135995 138009: contig of 2015 bp in length
* 138010 138109: gap of unknown length
* 138110 140502: contig of 2393 bp in length
* 140503 140602: gap of unknown length
* 140603 142745: contig of 2143 bp in length
* 142746 142845: gap of unknown length
* 142846 145293: contig of 2448 bp in length
* 145294 145393: gap of unknown length
* 145394 148334: contig of 2941 bp in length
* 148335 148434: gap of unknown length
* 148435 150495: contig of 2061 bp in length.
Location/Qualifiers
source
1..150495
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-265D21"

```

```

BASE COUNT 40913 a 32705 c 32685 g 40987 t 3205 others
ORIGIN

```

```

Query Match 1.3%: Score 21; DB 2: Length 150495;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1575 GATTCAGCTCTGCTCTTCA 1595
|||||
Db 129028 GATTCAGCTCTGCTCTTCA 129048

```

```

RESULT 19
AL513016 150720 bp DNA linear PRI 02-NOV-2001
LOCUS Human DNA sequence from clone Rp11-474021 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL513016
VERSION AL513016.5 GI:16605693
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 150720)
AUTHORS Ramsay, H.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hinxton@sanger.ac.uk

```

On Nov 3, 2001 this sequence version replaced gi:16501152.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Mp., WormPEP. Information on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/BCP/Chr1>
 Rp11-474021 is from the library RPEC1-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone Rp11-474021 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true left end of clone Rp11-474021 is at 1 in this sequence. The true left end of clone Rp5-1198H6 is at 148721 in this sequence.

```

FEATURES
source
1..150720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-474021"
/clone_lib="RPEC1-11.2"
/clone_lib="RPEC1-11.2"

```

misc_feature

```

BASE COUNT 40472 a 33567 c 34757 g 41924 t
ORIGIN

```

```

Query Match 1.3%: Score 21; DB 9: Length 150720;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 353 CTTTAAATTCTTGCCCT 373
|||||
Db 31 CTTTAAATTCTTGCCCT 11

```

```

RESULT 20
AC101941 158599 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone Rp24-112F21, WORKING DRAFT SEQUENCE, 23
DEFINITION unordered pieces.
ACCESSION AC101941
VERSION AC101941.2 GI:22381377
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 158599)
AUTHORS Birren, B., Nussbaum, C., and Lander, E.
TITLE Unpublished
JOURNAL 2 (bases 1 to 158599)

```

Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Collange, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
 Glinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Hulme, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Milhova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, C., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 158599)

Biren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,
Canarata, J., Chang, J., Chararo, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardya, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Li, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Milhova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced g1:117060717.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIDR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L17769
Center clone name: 112.F.21

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150624 bases at least Q40
Consensus quality: 154269 bases at least Q30
Consensus quality: 155666 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 156399; sum-of-ctrls
Quality coverage: 6.0 in Q20 bases; agarose-fp
Quality coverage: 5.7 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently
consists of 23 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 13026: contig of 13026 bp in length
* 13027 13126: gap of 100 bp
* 13127 13831: contig of 705 bp in length
* 13832 13931: gap of 100 bp

13932 14601: contig of 670 bp in length
* 14602 14701: gap of 100 bp
* 14702 15451: contig of 750 bp in length
* 15452 15551: gap of 100 bp
* 15552 16678: contig of 1127 bp in length
* 16679 16778: gap of 100 bp
* 16779 17604: contig of 826 bp in length
* 17605 17704: gap of 100 bp
* 17705 18923: contig of 1219 bp in length
* 18924 19023: gap of 100 bp
* 19024 20310: contig of 1287 bp in length
* 20311 20410: gap of 100 bp
* 20411 22661: contig of 2251 bp in length
* 22662 22761: gap of 100 bp
* 22762 23949: contig of 1188 bp in length
* 23950 24049: gap of 100 bp
* 24050 50090: contig of 26041 bp in length
* 50091 50190: gap of 100 bp
* 50191 51831: contig of 1641 bp in length
* 51832 51931: gap of 100 bp
* 51932 55850: contig of 3919 bp in length
* 55851 55950: gap of 100 bp
* 55951 59837: contig of 3887 bp in length
* 59838 59937: gap of 100 bp
* 59938 63375: contig of 3438 bp in length
* 63376 63475: gap of 100 bp
* 63476 71452: contig of 7977 bp in length
* 71453 71552: gap of 100 bp
* 71553 80301: contig of 8749 bp in length
* 80302 80401: gap of 100 bp
* 80402 90305: contig of 9904 bp in length
* 90306 90405: gap of 100 bp
* 90406 100251: contig of 9846 bp in length
* 100252 100351: gap of 100 bp
* 100352 113788: contig of 13437 bp in length
* 113789 113888: gap of 100 bp
* 113889 134105: contig of 20217 bp in length
* 134106 134205: gap of 100 bp
* 134206 157288: contig of 23083 bp in length
* 157289 157388: gap of 100 bp
* 157389 158599: contig of 1211 bp in length.

FEATURES
SOURCE
Location/Qualifiers

1..158599
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-112F21"
/clone_lib="RPCI-24 Male Mouse BAC"
1..13026
/note="assembly_fragment
clone_end:SP6
vector_side:left"
13127..13831
/note="assembly_fragment"
13932..14601
/note="assembly_fragment"
14702..15451
/note="assembly_fragment"
15552..16678
/note="assembly_fragment"
16779..17604
/note="assembly_fragment"
17705..18923
/note="assembly_fragment"
19024..20310
/note="assembly_fragment"
20411..22661
/note="assembly_fragment"
22762..23949
/note="assembly_fragment"
24050..50090
/note="assembly_fragment"
50191..51831
/note="assembly_fragment"

* 16486 18835: contig of 2350 bp in length
* 18836 18935: gap of unknown length
* 18936 21757: contig of 2822 bp in length
* 21758 21857: gap of unknown length
* 21858 24748: contig of 2891 bp in length
* 24749 24848: gap of unknown length
* 24849 27565: contig of 2717 bp in length
* 27566 30397: contig of 2732 bp in length
* 30398 30498: gap of unknown length
* 30498 34133: contig of 3636 bp in length
* 34134 34233: gap of unknown length
* 34234 36775: contig of 2342 bp in length
* 36776 36875: gap of unknown length
* 36876 40849: contig of 3974 bp in length
* 40850 40949: gap of unknown length
* 40950 44518: contig of 3569 bp in length
* 44519 44618: gap of unknown length
* 44619 48721: contig of 4103 bp in length
* 48722 48821: gap of unknown length
* 48822 52408: contig of 3587 bp in length
* 52409 52508: gap of unknown length
* 52509 55637: contig of 3129 bp in length
* 55638 55737: gap of unknown length
* 55738 59567: contig of 3830 bp in length
* 59568 59667: gap of unknown length
* 59668 62468: contig of 2801 bp in length
* 62469 62568: gap of unknown length
* 62569 66315: contig of 3747 bp in length
* 66316 66415: gap of unknown length
* 66416 70295: contig of 3860 bp in length
* 70296 70395: gap of unknown length
* 70396 75318: contig of 4923 bp in length
* 75319 75418: gap of unknown length
* 75419 81215: contig of 5797 bp in length
* 81216 81315: gap of unknown length
* 81316 86210: contig of 4895 bp in length
* 86211 86310: gap of unknown length
* 86311 92154: contig of 5844 bp in length
* 92155 92254: gap of unknown length
* 92255 97004: contig of 4750 bp in length
* 97005 97104: gap of unknown length
* 97105 103480: contig of 6376 bp in length
* 103481 103580: gap of unknown length
* 103581 108040: contig of 4460 bp in length
* 108041 108140: gap of unknown length
* 108141 113761: contig of 5621 bp in length
* 113762 113861: gap of unknown length
* 113862 118463: contig of 4602 bp in length
* 118464 118563: gap of unknown length
* 118564 127494: contig of 8931 bp in length
* 127495 127594: gap of unknown length
* 127595 134922: contig of 7328 bp in length
* 134923 135022: gap of unknown length
* 135023 144536: contig of 9514 bp in length
* 144537 144636: gap of unknown length
* 144637 155238: contig of 10602 bp in length
* 155239 155338: gap of unknown length
* 155339 165679: contig of 10341 bp in length.
* 165679 Location/Qualifiers
1.165679
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-402010"

BASE COUNT 50332 a 29730 c 29146 g 47927 t 8544 others
ORIGIN

Query Match 1.3%; Score 21; DB 2; Length 165679;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1576 ATTGAGCTCTGCTTCTCAT 1596
|||||

Db 53196 ATTCAGCTCTGCTTCTCAT 53216

RESULT 22

AP001554/c

LOCUS

AP001554 171747 bp DNA linear PRI 06-MAR-2001

DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-35p15,

complete sequence.

ACCESSION

AP001554

VERSION

AP001554.5 GI:13359361

KEYWORDS

HTG.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Query Match 1.3%; Score 21; DB 9; Length 171747;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 446 CCTACCTCAGATTGCTAGG 466
|||||
Db 94200 CCTACTCCAGATTGCTAGG 94180
RESULT 23
AC020766 173595 bp DNA linear PRI 23-AUG-2002
LOCUS Homo sapiens chromosome 16 clone RP11-70N16, complete sequence.
AC020766
ACCESSION AC020766
VERSION AC020766.9 GI:22450569
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE: 1 (bases 1 to 173595)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 173595)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 173595)
DOE Joint Genome Institute.
Direct Submission
Submitted (21-AUG-2002) Production Sequencing Facility, DOE Joint


```

* 79941 80040: gap of 100 bp
* 80041 91399: contig of 11359 bp in length
* 91400 91499: gap of 100 bp
* 91500 102181: contig of 10682 bp in length
* 102182 102281: gap of 100 bp
* 102282 114041: contig of 11760 bp in length
* 114042 114141: gap of 100 bp
* 114142 125232: contig of 11091 bp in length
* 125233 125332: gap of 100 bp
* 125333 132700: contig of 7368 bp in length
* 132701 132800: gap of 100 bp
* 132801 140074: contig of 7274 bp in length
* 140075 140174: gap of 100 bp
* 140175 145154: contig of 4980 bp in length
* 145155 145254: gap of 100 bp
* 145255 150197: contig of 4943 bp in length
* 150198 150297: gap of 100 bp
* 150298 154284: contig of 3987 bp in length
* 154285 154384: gap of 100 bp
* 154385 158901: contig of 4517 bp in length
* 158902 159001: gap of 100 bp
* 159002 162569: contig of 3568 bp in length
* 162570 162669: gap of 100 bp
* 162670 166056: contig of 3387 bp in length
* 166057 166156: gap of 100 bp
* 166157 169174: contig of 3018 bp in length
* 169175 169274: gap of 100 bp
* 169275 171034: contig of 1760 bp in length
* 171035 171134: gap of 100 bp
* 171135 173192: contig of 2058 bp in length
* 173193 173292: gap of 100 bp
* 173293 175034: contig of 1742 bp in length
* 175035 175134: gap of 100 bp
* 175135 176781: contig of 1647 bp in length
* 176782 176881: gap of 100 bp
* 176882 178330: contig of 1449 bp in length
* 178331 178430: gap of 100 bp
* 178431 180012: contig of 1582 bp in length
* 180013 180112: gap of 100 bp
* 180113 181489: contig of 1377 bp in length
* 181490 181589: gap of 100 bp
* 181590 183286: contig of 1699 bp in length
* 183289 183388: gap of 100 bp
* 183389 184908: contig of 1520 bp in length
* 184909 185008: gap of 100 bp
* 185009 186429: contig of 1421 bp in length
* 186430 186529: gap of 100 bp
* 186530 187961: contig of 1432 bp in length
* 187962 188061: gap of 100 bp
* 188062 189626: contig of 1565 bp in length
* 189627 189726: gap of 100 bp
* 189727 191063: contig of 1337 bp in length
* 191064 191163: gap of 100 bp
* 191164 192407: contig of 1244 bp in length
* 192408 192507: gap of 100 bp
* 192508 193854: contig of 1347 bp in length
* 193855 193954: gap of 100 bp
* 193955 195326: contig of 1372 bp in length
* 195327 195426: gap of 100 bp
* 195427 196485: contig of 1059 bp in length
* 196486 196585: gap of 100 bp
* 196586 197657: contig of 1072 bp in length
* 197658 197757: gap of 100 bp
* 197758 199077: contig of 1320 bp in length
* 199078 199177: gap of 100 bp
* 199178 200305: contig of 1128 bp in length
* 200306 200405: gap of 100 bp
* 200406 201591: contig of 1186 bp in length
* 201592 201691: gap of 100 bp
* 201692 202867: contig of 1176 bp in length
* 202868 202967: gap of 100 bp
* 202968 204235: contig of 1268 bp in length
* 204236 204335: gap of 100 bp

```

```

* 204336 205348: contig of 1013 bp in length
* 205349 205448: gap of 100 bp
* 205449 206463: contig of 1015 bp in length.
FEATURES
  source
    Location/Qualifiers
      . 206463
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="11"
        /map="11q23"
        /clone="RP11-853A16"
        . 34031
        /note="assembly_fragment"
        34132. .48505
        /note="assembly_fragment"
        48606. .62211
        /note="assembly_fragment"
        62312. .79940
        /note="assembly_fragment"

Query Match
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 446 CTTACCTCCAGGATTGCTTAG 466
Db 21073 CTTACCTCCAGGATTGCTTAG 21093

RESULT 25
AX098422
LOCUS
DEFINITION
  AX098422 Sequence 1 from Patent WO0120009.
ACCESSION
  AX098422
VERSION
  AX098422.1 GI:13537714
KEYWORDS
  SOURCE
    thale cress.
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
    1 (bases 1 to 1770)
    Reindl,A., Geigenberger,P.L., Neuhaus,H.E., Graeve-Kampfenkel,K.,
    Moehlmann,T. and Tjaden,J.
    Plants having altered amino acid contents and method for the
    production thereof
    Patent: WO 0120009-A 1 22-MAR-2001;
    JOURNAL
      BASF AKTIENGESELLSCHAFT (DE)
      Location/Qualifiers
        1. 1770
        /organism="Arabidopsis thaliana"
        /db_xref="taxon:3702"
        . 441 a 383 c 431 g 515 t
BASE COUNT
  ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1308 AAGAGAAATGGCTATATCC 1327
Db 1489 AAGAGAAATGGCTATATCC 1508

RESULT 26
AF428316
LOCUS
DEFINITION
  Arabidopsis thaliana At1g80300/F516_5 mRNA, complete cds.
ACCESSION
  AF428316
VERSION
  AF428316.1 GI:16226733
KEYWORDS
  FLI CDNA.
  Arabidopsis thaliana.
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

REFERENCE
AUTHORS

Spermatophyta: Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 2110)

TITLE
JOURNAL
REFERENCE
AUTHORS

Chen, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamuya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE
JOURNAL

Arabidopsis cdna clones
Unpublished
2 (bases 1 to 2110)
Chen, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamuya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

COMMENT

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Chen, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Shin, P., Banh, J., Bowser, L., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Chen, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES
SOURCE5'UTR
CDS3'UTR
BASE COUNT

1931 . 2110
525 a 454 c 502 g 629 t

ORIGIN

Query Match 1.2%; Score 20; DB 8; Length 2110;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

cy 1308 AAGCAATGCGCTATATGCC 1327
|||||
db 1553 AAGCAATGCGCTATATGCC 1572

RESULT 27

AY045903

LOCUS

DEFINITION Arabidopsis thaliana putative adenine nucleotide translocase (Atlg80300) mRNA, complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

Arabidopsis full length cDNA clones
Unpublished
2 (bases 1 to 2151)
Yamada, K., Liu, S.X., Pham, P.K., Banh, J., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamuya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shin, P., Southwick, A., Tracy, S.E., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

JOURNAL

COMMENT

Submitted (10-JUL-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamuya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X., Pham, P.K., Banh, J., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shin, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

FEATURES

SOURCE

Location/Qualifiers
1. 2151
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="RAFL07-08-J19 (R10784)"
/note="This clone is in a modified pBluescript vector (PIC-1) as a BamHI/XhoI insert.
ecotype: Columbia"

```

gene
1. .2151
/gene="Atlg80300"
5'UTR
1.27
/gene="Atlg80300"
CDS
28. .1902
/gene="Atlg80300"
/codon_start=1
/evidence=experimental
/product="putative adenine nucleotide translocase"
/protein_id="AKK76577.1"
/db_xref="GI:15028093"
/translation="MEAVIOTRGLSLPTKPIGVRSQLOPSHGLKQRLFAAKPRNLHG
LSLSENGHKKOTPEPTLHGISISHKERSTFICKAAAGDGAAPGEGSAAYVAS
PKIRGEVATLTKIIPGLMPCILFNTIILDTQDVLVYAKSSAEIIPPLKTVNL
LPMAGFMULTKLSNVLSKALFTVIVPTITGAFGVWPLPSNTIHPALADKL
LTUGPRMGPPLAIRIWSFCLFYVMAELGWSVSVLFGFANOITTVDEKKRPL
FGLGNVALIFSGRIVKYSNLKRLNGVDGMAVSLKAMSIYVGMICLILVMY
NRYVPLPTRSNKKKPKMGKMSLKLPLVSSPIYIDLATLVAVGISINLVEVTKSK
LKAQPSPEXSAFMGDFSTCGVATFTVMLLSIVFNKYGVAAKITTPVILITGV
AFPSILIRGSPAPLAKIGMTPLLAAYVGAALONIFSKAKYSLFDECKEMAYIPLD
EDTVYKKAIDVVCNPLGKSGALIQOFPMLISFSLANSIPYLGIMILLVITAMLA
AKSLGQPNLSRSEEELEKEMERASSVKIPVVSQDESGSLGESSSPSKSAFTNL"

misc_difference
869
/gene="Atlg80300"
/note="compared to genomic sequence resulting in an amino
acid sequence difference"
/replace="C"
1903. .2151
/gene="Atlg80300"
/misc_difference
2134
/gene="Atlg80300"
/note="compared to genomic sequence"
/replace="C"

BASE COUNT
553 a 451 c 510 g 637 t

ORIGIN
Query Match 1.2% Score 20; DB 8; Length 2151;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1308 AAAGAAATGGCCTATATCCC 1327
|||||
Db 1525 AAAGAAATGGCCTATATCCC 1544

RESULT 28
ATATMR 2181 bp mRNA linear PLN 26-NOV-1999
LOCUS A.thaliana mRNA for adenine nucleotide translocase.
DEFINITION Z49227
ACCESSION Z49227.2 GI:6469339
VERSION adenine nucleotide translocase.
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2181)
REFERENCE 1 Kampfenkel, K., Mohlmann, T., Batz, O., Van Montagu, M., Inze, D. and
AUTHORS Neuhaus, H.E.
TITLE Molecular characterization of an Arabidopsis thaliana cDNA encoding
JOURNAL a novel putative adenylate translocator of higher plants
MEDLINE FEBS Lett. 374 (3), 351-355 (1995)
PUBMED 96069943
REMARK 7589569
PUBMED (sites)
REFERENCE 2 (bases 1 to 2181)
AUTHORS Kampfenkel, K.K.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1995) Kampfenkel K.K., Universiteit Gent,
Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium,
B-9000 Gent

```

```

REMARK revised by [3]
REFERENCE 3 (bases 1 to 2181)
AUTHORS Kampfenkel, K.K.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1999) Kampfenkel K.K., Universiteit Gent,
Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium,
B-9000 Gent
COMMENT On Nov 27, 1999 this sequence version replaced gi:1051108.
FEATURES
source
1. .2181
/location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="var. Columbia"
/db_xref="taxon:3702"
/cloned="C23"
/issue_type="whole seedling"
/clone_1ib="lambdages"
/dev_lig="seedling"
110. .1981
/codon_start=1
/product="adenine nucleotide translocase"
/protein_id="CA89201.2"
/db_xref="GI:6469340"
/db_xref="SPTREMBL:Q39002"
/translation="MEAVIOTRGLSLPTKPIGVRSQLOPSHGLKQRLFAAKPRNLHG
CLYPLTGRNRPUSQPCMGFRPPTKRAPSSVARRRGRCWRSCLRSDSAAVASR
KIFGEVATLTKIIPGLMPCILFNTIILDTQDVLVYAKSSAEIIPPLKTVNL
LPMAGFMULTKLSNVLSKALFTVIVPTITGAFGVWPLPSNTIHPALADKL
LTUGPRMGPPLAIRIWSFCLFYVMAELGWSVSVLFGFANOITTVDEKKRPL
FGLGNVALIFSGRIVKYSNLKRLNGVDGMAVSLKAMSIYVGMICLILVMY
NRYVPLPTRSNKKKPKMGKMSLKLPLVSSPIYIDLATLVAVGISINLVEVTKSK
LKAQPSPEXSAFMGDFSTCGVATFTVMLLSIVFNKYGVAAKITTPVILITGV
AFPSILIRGSPAPLAKIGMTPLLAAYVGAALONIFSKAKYSLFDECKEMAYIPLD
EDTVYKKAIDVVCNPLGKSGALIQOFPMLISFSLANSIPYLGIMILLVITAMLA
AKSLGQPNLSRSEEELEKEMERASSVKIPVVSQDESGSLGESSSPSKSAFTNL"

BASE COUNT
549 a 476 c 508 g 648 t

ORIGIN
Query Match 1.2% Score 20; DB 8; Length 2181;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1308 AAAGAAATGGCCTATATCCC 1327
|||||
Db 1604 AAAGAAATGGCCTATATCCC 1623

RESULT 29
AF222985
LOCUS AF222985 2762 bp DNA linear PRI 02-JUN-2000
DEFINITION Homo sapiens derived chromosome 1 translocation breakpoint
sequence.
ACCESSION AF222985
VERSION AF222985.1 GI:8163838
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 2762)
AUTHORS Miller, J.K., Wilson-Annan, J.C., Anderson, S., Christie, S.,
Taylor, M.S., Sempile, C.A., Devon, R.S., Clair, D.M., Muir, W.J.,
Blackwood, D.H. and Porteous, D.J.
TITLE Disruption of two novel genes by a translocation co-segregating
JOURNAL Hum. Mol. Genet. 9 (9), 1415-1423 (2000)
MEDLINE 20275630
PUBMED 10814723
REFERENCE 2 (bases 1 to 2762)
AUTHORS Miller, K., Anderson, S. and Christie, S.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Medical Genetics Section, The University of
Edinburgh, Crewe Road, Edinburgh EH4 2XU, Scotland
FEATURES
Location/Qualifiers

```

source	1. .2762	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/map="1;11) (q42.1;q14.3)"	
	/note="the derived 1 chromosome results from a balanced translocation between chromosomes 1 and 11; DNA has been lost from 1q42.1-qter and replaced with chromosome 11 material from 11q14.3-qter."	
source	1. .1663	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/chromosome="1"	
	/map="1q42.1"	
source	1664. .2762	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/chromosome="11"	
	/map="11q14.3"	
BASE COUNT	749 a 570 c 511 g 932 t	
ORIGIN		
Query Match	1.2%; Score 20; DB 9; Length 2762;	
Best Local Similarity	100.0%; Pred. No. 37;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	913 TTTTATTACTCTCTTGCTT 932	
Db	339 TTTTATTACTCTCTTGCTT 358	
RESULT 30		
AF222983		
LOCUS	7291 bp DNA linear PRI 02-JUN-2000	
DEFINITION	Homo sapiens DISC1 protein (DISC1) gene, partial cds and DISC2	
ACCESSION	AF222983	
VERSION	AF222983	
KEYWORDS	AF222983.1 GI:8163835	
SOURCE		
ORGANISM	Homo sapiens.	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 7291)	
AUTHORS	Millar,J.K., Wilson-Annan,J.C., Anderson,S., Christie,S., Taylor,M.S., Semple,C.A., Devon,R.S., Claff,D.M., Muir,W.J., Blackwood,D.H. and Porteous,D.J.	
TITLE	Disruption of two novel genes by a translocation co-segregating with Schizophrenia	
JOURNAL	Hum. Mol. Genet. 9 (9), 1415-1423 (2000)	
MEDLINE	20275630	
PUBMED	10814723	
REFERENCE	2 (bases 1 to 7291)	
AUTHORS	Millar,K., Anderson,S. and Christie,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-JUN-2000) Medical Genetics Section, The University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, Scotland	
FEATURES	Location/Qualifiers	
source	1. .7291	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/chromosome="1"	
	/map="1q42.1"	
	complement(<1. .>7291)	
gene	/gene="DISC2"	
misc_feature	1. .7291	/note="translocation breakpoint flanking region"
exon	complement(1. .7291)	
	/gene="DISC2"	
	/note="non coding"	
gene	<5357. .>5545	
	/gene="DISC1"	
	<5357. .>5545	
mRNA	/gene="DISC1"	

```

CDS
<5357..>5545
/gene="DISC1"
/codon_start=3
/product="DISC1 protein"
/protein_id="AA073874.1"
/db_xref="GI:8163836"
/translation="NHFWPAKDLTEFIRSLTSENEGELGLSKLVLSRRNKKLGSV
KEDINRIRREVEHOETAY"
KEDINRIRREVEHOETAY"
<5357..>5545
/gene="DISC1"
/number=9

BASE COUNT      1968 a      1354 c      1430 g      2526 t      13 others
ORIGIN

Query Match      1.2%; Score 20; DB 9; Length 7291;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      913 TTTTATTAACTCTCTGTGGTT 932
        |||||||
Db      339 TTTTATTAACTCTCTGTGGTT 358

RESULT 31
AF222981/c
LOCUS      AF222981      15002 bp      DNA      linear      PRI 02-JUN-2000
DEFINITION Homo sapiens 1 DISC2 gene, complete sequence.
ACCESSION  AF222981
VERSION     AF222981.1  GI:8163870
KEYWORDS
SOURCE      Homo sapiens.
            Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 15002)
            Millar,J.K., Wilson-Annan,J.C., Anderson,S., Christie,S.,
            Taylor,M.S., Sempile,C.A., Devon,R.S., Clair,D.M., Muir,W.J.,
            Blackwood,D.H. and Porteous,D.J.
            Disruption of two novel genes by a translocation co-segregating
            with schizophrenia
            Hum. Mol. Genet. 9 (9), 1415-1423 (2000)
JOURNAL     20275630
MEDLINE     10614723
PUBMED      2 (bases 1 to 15002)
REFERENCE   Millar,K., Anderson,S. and Christie,S.
            Direct Submission
JOURNAL     Submitted (11-JUN-2000) Medical Genetics Section, The University of
            Edinburgh, Crewe Road, Edinburgh EH4 2XU, Scotland
FEATURES
            source
            1..15002
                Location/Qualifiers
                1..15002
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="1"
                    /map="1q42.1"
                    1..15002
                        /gene="DISC2"
                        /note="considered to be a non-coding; 5' end has not yet
                        been identified"
gene
    .
BASE COUNT      4815 a      3132 c      2838 g      4209 t      8 others
ORIGIN

Query Match      1.2%; Score 20; DB 9; Length 15002;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      913 TTTTATTAACTCTCTGTGGTT 932
        |||||||
Db      11647 TTTTATTAACTCTCTGTGGTT 11628

RESULT 32
AF222987
LOCUS      AF222987      33376 bp      DNA      linear      PRI 02-JUN-2000

```

DEFINITION Homo sapiens DISC1 protein (DISC1) gene, partial cds and DISC2
gene, partial sequence.
ACCESSION AF222987
VERSION AF222987.1 GI:8163844
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 33376)
AUTHORS Miller,J.K., Wilson-Annan,J.C., Anderson,S., Christie,S.,
Taylor,M.S., Semple,C.A., Devon,R.S., Clair,D.M., Mulr,W.J.,
Blackwood,D.H. and Porteous,D.J.
TITLE Disruption of two novel genes by a translocation co-segregating
with schizophrenia
JOURNAL Hum. Mol. Genet. 9 (9), 1415-1423 (2000)
MEDLINE 20275630
PUBMED 10814723
REFERENCE 2 (bases 1 to 33376)
AUTHORS Anderson,S., Miller,K. and Christie,S.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2000) Medical Genetics Section, The University of
Edinburgh, Crewe Road, Edinburgh EH4 2XU, Scotland
FEATURES
Source Location/Qualifiers
1..33376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q42.1"
1..33376
/note="Translocation breakpoint region"
1..8206
/note="contig1"
<4771..>26773
/gene="DISC1"
join(<4771..4825,8413..8515,26585..>26773)
/gene="DISC1"
/codon_start=2
/product="DISC1 protein"
/protein_id="AAF73877.1"
/db_xref="GI:8163845"
/translation="LQBRKISNLSKEITTKVMSRFCSSTLRKKNDETQTPALL
EAKKHAISGNHFWAKDLTEIRINSLSREGGLSKLLVLSRNKKLGSKVEDYN
RLRVEVHETAY"
4771..4825
/gene="DISC1"
/number=7
8206..33376
/note="contig2"
8413..8515
/gene="DISC1"
/number=8
complement(18199..>33376)
/gene="DISC2"
complement(18199..>33376)
/gene="DISC2"
26585..26773
/gene="DISC1"
/number=9
BASE COUNT 9071 a 6789 c 7155 g 10350 t 11 others
ORIGIN
Query Match 1.2%: Score 20; DB 9; Length 33376;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 913 TTTTATTACTCTCTGGTT 932
|||||
Db 21554 TTTTATTACTCTCTGGTT 21573
RESULT 33
AX151213

LOCUS AX151213 33780 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 4 from Patent WO0140301.
ACCESSION AX151213
VERSION AX151213.1 GI:14533379
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 33780)
AUTHORS Porteous,D., Miller,K. and Blackwood,D.
TITLE A gene, disrupted in schizophrenia
JOURNAL Patent: WO 0140301-A 4 07-JUN-2001;
Akzo Nobel N.V. (NL); MEDICAL RESEARCH COUNCIL (GB); UNIVERSITY
OF EDINBURGH (GB)
FEATURES
Source Location/Qualifiers
1..33780
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 9140 a 6838 c 7205 g 10408 t 189 others
ORIGIN
Query Match 1.2%: Score 20; DB 6; Length 33780;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 913 TTTTATTACTCTCTGGTT 932
|||||
Db 21958 TTTTATTACTCTCTGGTT 21977
RESULT 34
AC107059 48546 bp DNA linear PRI 20-APR-2002
AC107059
LOCUS
DEFINITION Homo sapiens BAC clone RP11-668G10 from 4, complete sequence.
ACCESSION AC107059
VERSION AC107059.6 GI:19848496
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 48546)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 48546)
AUTHORS Cedroni,M. and Cotton,M.
TITLE The sequence of Homo sapiens BAC clone RP11-668G10
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 48546)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 48546)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 48546)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 48546)
AUTHORS Waterston,R.
TITLE Direct Submission

JOURNAL

Submitted (20-APR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 30, 2002 this sequence version replaced gi:19551219.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0668610

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frenken, E., Tatenno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-336N6, 2000 bp overlap; the clone sequenced to the right is RP11-469011, 2000 bp overlap.

Actual start of this clone is at base position 123490 of RP11-336N6; actual end is at base position 104312 of RP11-469011.

Polymorphisms have been identified between AC107059, AC055120 and AC012504.

FEATURES

SOURCE

1. 48546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-668610"
/clone_lib="RPCI-11"
112..127
/note="match to EST AI654124 (NID:g4738103) ty62a05.x1"
112..125
/note="similar to Homo sapiens EST AI741715 (NID:g5110003) wg22b05.x1"
112..123
/note="match to EST AW449862 (NID:g6990638)"
113..124
/note="match to EST AI962124 (NID:g5754837) wx77f01.x1"
119..438
/note="similar to Homo sapiens EST HT8148 (NID:g1056237) yu86ff10.r1"
122..385
/note="similar to Homo sapiens EST AA640399 (NID:g2565649) nt99e01.s1"
misc_feature
126..812

/note="similar to Homo sapiens EST BF669641 (NID:g11943536)"
342..352
/note="match to EST AI962124 (NID:g5754837) wx77f01.x1"
393..749
/note="match to EST AI969703 (NID:g5766521) wz70f08.x1"
450..848
/note="match to EST AA863006 (NID:g2955485) og99d10.s1"
546..952
/note="similar to EST BM541231 (NID:g18824674)"
598..1112
/note="similar to Homo sapiens EST AI741715 (NID:g5110003) wg22b05.x1"
600..900
/note="match to EST AA401689 (NID:g2057173) zt66a07.s1"
658..1140
/note="similar to Homo sapiens EST AU525280 (NID:g12788773)"
671..1347
/note="similar to Homo sapiens EST B1490709 (NID:g15329937)"
674..1364
/note="similar to Homo sapiens EST AU525322 (NID:g12788815)"
690..1350
/note="similar to Homo sapiens EST AU546207 (NID:g12879099)"
719..1190
/note="similar to Homo sapiens EST AU574646 (NID:g12935046)"
724..1130
/note="similar to Homo sapiens EST AA283104 (NID:g1926029) zt16a12.r1"
967..968
/note="similar to Homo sapiens EST N64351 (NID:g1212180) yz47e12.s1"
1007..1175
/note="similar to Homo sapiens EST N64351 (NID:g1212180) yz47e12.s1"
1017..1362
/note="match to EST AL036737 (NID:g5927885)"
1017..1362
/note="similar to Homo sapiens EST ALJ35195 (NID:g6603382)"
1074..1350
/note="similar to EST BM671395 (NID:g18981293)"
1135..1681
/note="similar to EST BM678317 (NID:g18988213)"
1145
/note="similar to Homo sapiens EST AI741715 (NID:g5110003) wg22b05.x1"
1206..1356
/note="similar to Homo sapiens EST AU525280 (NID:g12788773)"
1350..1387
/rpt_family="(CA)n"
1380..1498
/note="similar to Homo sapiens EST AU525280 (NID:g12788773)"
1385..1603
/note="match to EST BG675536 (NID:g13906932)"
1386..1595
/note="match to EST AL036737 (NID:g5927885)"
1386..1561
/note="similar to Homo sapiens EST AU525322 (NID:g12788815)"
1386..1438
/note="similar to Homo sapiens EST ALJ35195 (NID:g6603382)"
1396..1595
/note="match to EST AW515614 (NID:g7153696) hd77d06.x1"
1397..1981
/note="match to EST AW007714 (NID:g5856492) wt68c03.x1"

```

misc_feature      1403..1595
                  /note="match to EST A1654124 (NID:g4738103) ty62a06.x1"
misc_feature      1404..1595
                  /note="similar to Mus musculus EST BF152508
                  (NID:g11033903) uz34b11.y1"
misc_feature      1414..1716
                  /note="similar to Homo sapiens EST AA406045 (NID:g2064073)
                  zu67d07.s1"
misc_feature      1422..1841
                  /note="match to EST AA292874 (NID:g1941855) zt66a07.r1"
misc_feature      1438..2084
                  /note="similar to Mus musculus EST AV258065
                  (NID:g16388469)"
misc_feature      1472..2203

```

```

Query Match      1.2%; Score 20; DB 9; Length 48546;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      19 AATGACAAATTAAGTATT 38
        |||||||
Db      12988 AATGACAAATTAAGTATT 13007

```

```

RESULT 35
ALJ34721/c      ALJ34721      57070 bp      DNA      linear      PRI 15-NOV-2001
DEFINITION      Human DNA sequence from clone RP11-373E22 on chromosome 1, complete
                  sequence.
ACCESSION      ALJ34721
KEYWORDS
SOURCE      HTG.
ORGANISM      human.
                Homo sapiens
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 57070)
AUTHORS      Mashreghi-Mohammadi, M.
TITLE      Direct Submission
JOURNAL
COMMENT      Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
                hummer@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
                On Nov 16, 2001 this sequence version replaced gi:16030110.
                During sequence assembly data is compared from overlapping clones.
                Where differences are found these are annotated as variations
                together with a note of the overlapping clone name. Note that the
                variation annotation may not be found in the sequence submission
                corresponding to the overlapping clone, as we submit sequences with
                only a small overlap as described above.
                This sequence was finished as follows unless otherwise noted: all
                regions were either double-stranded or sequenced with an alternate
                chemistry or covered by high quality data (i.e., phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats: all regions were covered by at least
                one plasmid subclone or more than one M13 subclone; and the
                assembly was confirmed by restriction digest. The following
                abbreviations are used to associate primary accession numbers given
                in the feature table with their source databases: Em:, EMBL; SW:,
                SWISSPROT; Tr:, TrEMBL; Wp:, WormPeP; Information on the WormPeP
                database can be found at
                http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
                was generated from part of bacterial clone contigs of human
                chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
                Group. Further information can be found at
                http://www.sanger.ac.uk/HGP/Chr1
                RP11-373E22 is from the library RPCT-11.2 constructed by the group
                of Pieter de Jong. For further details see
                http://www.chori.org/bacpac/home.htm
                VECTOR: pBAC3.6

```

IMPORTANT: This sequence is not the entire insert of clone RP11-373E22. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-373E22 is at 57070 in this

```

FEATURES
    source          location/qualifiers
                   1..57070
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="1"
                   /clone="RP11-373E22"
                   /clone_lib="RPCT-11.2"

```

```

BASE COUNT      19390 a 10751 c 10316 g 16613 t
ORIGIN

```

```

Query Match      1.2%; Score 20; DB 9; Length 57070;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      3 AATAAACATATCAGATA 22
        |||||||
Db      35649 AATAAACATATCAGATA 35630

```

```

RESULT 36
AC113541
LOCUS      AC113541      58810 bp      DNA      linear      HTG 15-MAY-2002
DEFINITION      Mus musculus clone RP23-185N2, LOW-PASS SEQUENCE SAMPLING.
ACCESSION      AC113541
VERSION      AC113541.2 GI:20800304
KEYWORDS      HTGS_PHASED.
SOURCE      Mus musculus.
                Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 58810)
AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE      Mus musculus, clone RP23-185N2
JOURNAL      Unpublished
AUTHORS

```

```

REFERENCE
TITLE      2 (bases 1 to 58810)
JOURNAL
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
                MacDonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M.,
                McKean, P., McKernan, K., Meldrum, J., Meus, L., Mihova, T.,
                Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
                Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
                Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
                Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
                Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
                Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
                Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
                Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
                Vriel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
                Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
                Direct Submission
                Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
                3 (bases 1 to 58810)
                Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
                Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
                Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
                Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
                Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
                Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
                Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
                Hagos, B., Horton, L., Huine, W., Iliev, I., Johnson, R., Jones, C.,
                Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,

```


TITLE
JOURNAL
COMMENT

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margus, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., Meldrum, J., Meneses, L., Milnova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Rella, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Rhmann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemlek, L., Zimmer, A. and Zody, M.

Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 15, 2002 this sequence version replaced g1:19033725.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L24218

Center clone name: 185_N_2

* NOTE: This record contains 72 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 738 837: gap of 100 bp
* 838 1574: contig of 737 bp in length
* 1575 1674: gap of 100 bp
* 1675 2407: contig of 733 bp in length
* 2408 2507: gap of 100 bp
* 2508 3219: contig of 712 bp in length
* 3220 3319: gap of 100 bp
* 3320 4016: contig of 697 bp in length
* 4017 4116: gap of 100 bp
* 4117 4828: contig of 712 bp in length
* 4829 4928: gap of 100 bp
* 4929 5645: contig of 717 bp in length
* 5646 5745: gap of 100 bp
* 5746 6454: contig of 709 bp in length
* 6455 6554: gap of 100 bp
* 6555 7287: contig of 733 bp in length
* 7288 7387: gap of 100 bp
* 7388 8122: contig of 735 bp in length
* 8123 8222: gap of 100 bp
* 8223 8955: contig of 733 bp in length
* 8956 9055: gap of 100 bp
* 9056 9765: contig of 710 bp in length
* 9766 9865: gap of 100 bp
* 9866 10570: contig of 705 bp in length
* 10571 10670: gap of 100 bp
* 10671 11393: contig of 723 bp in length
* 11394 11493: gap of 100 bp
* 11494 12213: contig of 720 bp in length
* 12214 12313: gap of 100 bp
* 12314 13018: contig of 705 bp in length
* 13019 13118: gap of 100 bp
* 13119 13826: contig of 708 bp in length

* 13827 13926: gap of 100 bp
* 13927 14556: contig of 730 bp in length
* 14557 14756: gap of 100 bp
* 14757 15465: contig of 709 bp in length
* 15466 15565: gap of 100 bp
* 15566 16301: contig of 736 bp in length
* 16302 16401: gap of 100 bp
* 16402 17137: contig of 736 bp in length
* 17138 17237: gap of 100 bp
* 17238 17950: contig of 713 bp in length
* 17951 18050: gap of 100 bp
* 18051 18772: contig of 722 bp in length
* 18773 18872: gap of 100 bp
* 18873 19569: contig of 697 bp in length
* 19570 19669: gap of 100 bp
* 19670 20409: contig of 740 bp in length
* 20410 20509: gap of 100 bp
* 20510 21216: contig of 707 bp in length
* 21217 21316: gap of 100 bp
* 21317 22044: contig of 728 bp in length
* 22045 22144: gap of 100 bp
* 22145 22882: contig of 738 bp in length
* 22883 22982: gap of 100 bp
* 22983 23712: contig of 732 bp in length
* 23713 23812: gap of 100 bp
* 23813 24530: contig of 718 bp in length
* 24531 24630: gap of 100 bp
* 24631 25349: contig of 719 bp in length
* 25350 25449: gap of 100 bp
* 25450 26166: contig of 717 bp in length
* 26167 26266: gap of 100 bp
* 26267 26976: contig of 710 bp in length
* 26977 27076: gap of 100 bp
* 27077 27803: contig of 727 bp in length
* 27804 27903: gap of 100 bp
* 27904 28613: contig of 710 bp in length
* 28614 28713: gap of 100 bp
* 28714 29416: contig of 703 bp in length
* 29417 29516: gap of 100 bp
* 29517 30245: contig of 729 bp in length
* 30246 30345: gap of 100 bp
* 30346 31088: contig of 743 bp in length
* 31089 31188: gap of 100 bp
* 31189 31923: contig of 735 bp in length
* 31924 32023: gap of 100 bp
* 32024 32729: contig of 706 bp in length
* 32730 32829: gap of 100 bp
* 32830 33538: contig of 709 bp in length
* 33539 33638: gap of 100 bp
* 33639 34345: contig of 707 bp in length
* 34346 34445: gap of 100 bp
* 34446 35161: contig of 716 bp in length
* 35162 35261: gap of 100 bp
* 35262 36000: contig of 739 bp in length
* 36001 36100: gap of 100 bp
* 36101 36805: contig of 705 bp in length
* 36806 36905: gap of 100 bp
* 36906 37643: contig of 738 bp in length
* 37644 37743: gap of 100 bp
* 37744 38479: contig of 736 bp in length
* 38480 38579: gap of 100 bp
* 38580 39303: contig of 724 bp in length
* 39304 39403: gap of 100 bp
* 39404 40130: contig of 727 bp in length
* 40131 40230: gap of 100 bp
* 40231 40908: contig of 678 bp in length
* 40909 41008: gap of 100 bp
* 41009 41728: contig of 720 bp in length
* 41729 41828: gap of 100 bp
* 41829 42534: contig of 706 bp in length
* 42535 42634: gap of 100 bp
* 42635 43333: contig of 699 bp in length
* 43334 43433: gap of 100 bp

* 16371 17070: contig of 700 bp in length
* 17071 17170: gap of 100 bp
* 17171 17886: contig of 716 bp in length
* 17887 17986: gap of 100 bp
* 17987 18696: contig of 710 bp in length
* 18697 18796: gap of 100 bp
* 18797 19516: contig of 720 bp in length
* 19517 19616: gap of 100 bp
* 19617 20334: contig of 718 bp in length
* 20335 20434: gap of 100 bp
* 20435 21160: contig of 726 bp in length
* 21161 21260: gap of 100 bp
* 21261 21993: contig of 733 bp in length
* 21994 22093: gap of 100 bp
* 22094 22840: contig of 747 bp in length
* 22841 22940: gap of 100 bp
* 22941 23664: contig of 724 bp in length
* 23665 23764: gap of 100 bp
* 23765 24448: contig of 684 bp in length
* 24449 24548: gap of 100 bp
* 24549 25266: contig of 718 bp in length
* 25267 25366: gap of 100 bp
* 25367 26078: contig of 712 bp in length
* 26079 26178: gap of 100 bp
* 26179 26891: contig of 713 bp in length
* 26892 26991: gap of 100 bp
* 26992 27684: contig of 693 bp in length
* 27685 27784: gap of 100 bp
* 27785 28489: contig of 705 bp in length
* 28490 28589: gap of 100 bp
* 28590 29303: contig of 714 bp in length
* 29304 29403: gap of 100 bp
* 29404 30139: contig of 736 bp in length
* 30140 30239: gap of 100 bp
* 30240 30960: contig of 721 bp in length
* 30961 31060: gap of 100 bp
* 31061 31764: contig of 704 bp in length
* 31765 31864: gap of 100 bp
* 31865 32583: contig of 719 bp in length
* 32584 32683: gap of 100 bp
* 32684 33394: contig of 711 bp in length
* 33395 33494: gap of 100 bp
* 33495 34213: contig of 719 bp in length
* 34214 34313: gap of 100 bp
* 34314 35030: contig of 717 bp in length
* 35031 35130: gap of 100 bp
* 35131 35844: contig of 714 bp in length
* 35845 35944: gap of 100 bp
* 35945 36669: contig of 725 bp in length
* 36670 36769: gap of 100 bp
* 36770 37495: contig of 726 bp in length
* 37496 37595: gap of 100 bp
* 37596 38320: contig of 725 bp in length
* 38321 38420: gap of 100 bp
* 38421 39162: contig of 742 bp in length
* 39163 39262: gap of 100 bp
* 39263 40009: contig of 747 bp in length
* 40010 40109: gap of 100 bp
* 40110 40821: gap of 712 bp in length
* 40822 40921: gap of 100 bp
* 40922 41627: contig of 706 bp in length
* 41628 41727: gap of 100 bp
* 41728 42437: contig of 710 bp in length
* 42438 42537: gap of 100 bp
* 42538 43250: contig of 713 bp in length
* 43251 43350: gap of 100 bp
* 43351 44066: contig of 716 bp in length
* 44067 44166: gap of 100 bp
* 44167 44882: contig of 716 bp in length
* 44883 44982: gap of 100 bp
* 44983 45702: contig of 720 bp in length

Query Match

1.2% Score 20: DB 2: Length 65259;

Best Local Similarity 100.0%: Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 GAGCTAATATTCCTTACTA 641
|||||
Db 25639 GAGCTAATATTCCTTACTA 25658

RESULT 38
AC114147/C
LOCUS
DEFINITION
Rattus norvegicus clone CH230-144p3, *** SEQUENCING IN PROGRESS
***, 45 unordered pieces.
AC114147.2 GI:21732759
AC114147
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 74010)
Muzny,D.M., Adams,C., Adjo-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devaney,K.R., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Dellanev,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.,H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Hornst,E., Howard,S., Huber,J., Huliyk,S., Hume,D., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtharge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Matinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peety,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,L.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,
Umsani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,D., Watlington,S.,
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 74010)
Worley,K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 74010)
Worley,K.C.
Direct Submission

JOURNAL

COMMENT

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:19224474.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GLOB

Center clone name: CH230-144F3

Summary Statistics

Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 24661 bases at least Q40
Consensus quality: 26589 bases at least Q30
Consensus quality: 27921 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      1381: contig of 1381 bp in length
*      1382      1481: gap of unknown length
*      1482      2791: contig of 1310 bp in length
*      2792      2891: gap of unknown length
*      2892      4184: contig of 1293 bp in length
*      4185      4284: gap of unknown length
*      4285      5398: contig of 1114 bp in length
*      5399      5498: gap of unknown length
*      5499      6596: contig of 1098 bp in length
*      6597      6696: gap of unknown length
*      6697      7726: contig of 1030 bp in length
*      7727      7826: gap of unknown length
*      7827      8916: contig of 1090 bp in length
*      8917      9016: gap of unknown length
*      9017      10122: contig of 1106 bp in length
*      10123      11222: gap of unknown length
*      11222      11326: contig of 1004 bp in length
*      11327      12773: contig of 1447 bp in length
*      12774      12873: gap of unknown length
*      12873      13949: contig of 1076 bp in length
*      13950      14049: gap of unknown length
*      14050      15641: contig of 1592 bp in length
*      15642      15741: gap of unknown length
*      15742      17427: contig of 1686 bp in length
*      17428      17527: gap of unknown length
*      17528      18919: contig of 1392 bp in length
*      18920      19019: gap of unknown length
*      19020      20050: contig of 1031 bp in length
*      20051      21271: contig of 1121 bp in length
*      21272      21371: gap of unknown length
*      21372      22617: contig of 1246 bp in length
*      22618      22717: gap of unknown length
*      22717      23733: contig of 1016 bp in length
*      23734      23833: gap of unknown length
*      23834      25251: contig of 1418 bp in length
*      25252      25351: gap of unknown length
*      25352      26502: contig of 1151 bp in length
*      26503      28122: contig of 1520 bp in length
*      28123      28222: gap of unknown length
*      28223      29582: contig of 1360 bp in length

```

```

*      29583      29682: gap of unknown length
*      29683      31194: contig of 1512 bp in length
*      31195      31294: gap of unknown length
*      31295      32524: contig of 1230 bp in length
*      32525      32624: gap of unknown length
*      32625      34487: contig of 1863 bp in length
*      34488      34587: gap of unknown length
*      34588      36096: contig of 1509 bp in length
*      36097      37881: gap of unknown length
*      37882      37981: gap of unknown length
*      37982      39421: contig of 1440 bp in length
*      39422      39521: gap of unknown length
*      39522      41762: contig of 2241 bp in length
*      41763      41862: gap of unknown length
*      41863      43073: contig of 1211 bp in length
*      43074      43173: gap of unknown length
*      43174      44522: contig of 1349 bp in length
*      44523      44622: gap of unknown length
*      44623      46552: contig of 1930 bp in length
*      46553      48103: gap of unknown length
*      48104      48203: contig of 1451 bp in length
*      48204      49906: gap of unknown length
*      49907      50006: contig of 1703 bp in length
*      50007      51069: contig of 1063 bp in length
*      51070      51169: gap of unknown length
*      51170      52193: contig of 1024 bp in length
*      52194      52293: gap of unknown length
*      52294      53689: contig of 1386 bp in length
*      53690      53789: gap of unknown length
*      53790      55812: contig of 2023 bp in length
*      55813      55912: gap of unknown length
*      55913      58145: contig of 2233 bp in length
*      58146      58245: gap of unknown length
*      58246      59827: contig of 1682 bp in length
*      59828      60027: gap of unknown length
*      60028      61826: contig of 1799 bp in length
*      61827      61926: gap of unknown length
*      61927      64162: contig of 2236 bp in length
*      64163      64262: gap of unknown length
*      64263      67752: contig of 3490 bp in length
*      67753      67852: gap of unknown length
*      67853      71791: contig of 3939 bp in length
*      71792      71891: gap of unknown length
*      71892      74010: contig of 2119 bp in length.

```

FEATURES

source 1. 74010
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-144F3"

BASE COUNT 17870 a 15598 c 17993 g 18093 t 4456 others

ORIGIN

Query Match 1.2%; Score 20; DB 2; Length 74010;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1482 TTCATCATTCGTAATTGGTT 1501
|||||

Db 73928 TTCATCATTCGTAATTGGTT 73909

RESULT 39

AC087717 74274 bp DNA linear HTG 16-JAN-2001

LOCUS AC087717

DEFINITION Homo sapiens chromosome 11 clone RP1-145M24 map 11, LOW-PASS

SEQUENCE SAMPLING.

ACCESSION AC087717

VERSION AC087717.1 GI:12232529

KEYWORDS HTG: HNGS, PHASEO.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 74274)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP1-145M24
Unpublished
2 (bases 1 to 74274)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Boguslavsky,L., Boukhaltier,B., Brown,A.,
Camataia,J., Campolano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardina,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., MacDonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Petersen,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
Sounguez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Travers,M., Travis,N., Trigliio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence-submissions@genome.wi.mit.edu
Project Information
Center project name: L12105
Center clone name: 145_M24

NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
* 751 850: contig of 750 bp in length
* 851 1637: contig of 787 bp in length
* 1638 1737: gap of 100 bp
* 1738 2478: contig of 741 bp in length
* 2479 2578: gap of 100 bp
* 2579 3294: contig of 716 bp in length
* 3295 3394: gap of 100 bp
* 3395 4106: contig of 712 bp in length
* 4107 4206: gap of 100 bp
* 4207 4886: contig of 680 bp in length
* 4887 4986: gap of 100 bp
* 4987 5679: contig of 693 bp in length
* 5680 5779: gap of 100 bp
* 5780 6507: contig of 728 bp in length
* 6508 6607: gap of 100 bp
* 6608 7351: contig of 744 bp in length
* 7352 7451: gap of 100 bp
* 7452 8211: contig of 760 bp in length

8212 8311: gap of 100 bp
8312 9026: contig of 715 bp in length
9027 9126: gap of 100 bp
9127 9881: contig of 755 bp in length
9882 9981: gap of 100 bp
9982 10716: contig of 735 bp in length
10717 10816: gap of 100 bp
10817 11535: contig of 719 bp in length
11536 11635: gap of 100 bp
11636 12299: contig of 664 bp in length
12300 12399: gap of 100 bp
12400 13160: contig of 761 bp in length
13161 13260: gap of 100 bp
13261 13998: contig of 738 bp in length
13999 14098: gap of 100 bp
14099 14831: contig of 733 bp in length
14832 14931: gap of 100 bp
14932 15680: contig of 749 bp in length
15681 15780: gap of 100 bp
15781 16526: contig of 746 bp in length
16527 16626: gap of 100 bp
16627 17394: contig of 768 bp in length
17395 17494: gap of 100 bp
17495 18212: contig of 718 bp in length
18213 18312: gap of 100 bp
18313 19052: contig of 740 bp in length
19053 19152: gap of 100 bp
19153 19912: contig of 760 bp in length
19913 20012: gap of 100 bp
20013 20740: contig of 728 bp in length
20741 20840: gap of 100 bp
20841 21566: contig of 726 bp in length
21567 21666: gap of 100 bp
21667 22334: contig of 668 bp in length
22335 22434: gap of 100 bp
22435 23181: contig of 747 bp in length
23182 23281: gap of 100 bp
23282 24020: contig of 739 bp in length
24021 24120: gap of 100 bp
24121 24855: contig of 735 bp in length
24856 24955: gap of 100 bp
24956 25709: contig of 754 bp in length
25710 25809: gap of 100 bp
25810 26550: contig of 741 bp in length
26551 26650: gap of 100 bp
26651 27392: contig of 742 bp in length
27393 27492: gap of 100 bp
27493 28232: contig of 740 bp in length
28233 28332: gap of 100 bp
28333 29064: contig of 732 bp in length
29065 29164: gap of 100 bp
29165 29884: contig of 720 bp in length
29885 29984: gap of 100 bp
29985 30676: contig of 692 bp in length
30677 30776: gap of 100 bp
30777 31527: contig of 751 bp in length
31528 31627: gap of 100 bp
31628 32362: contig of 735 bp in length
32363 32462: gap of 100 bp
32463 33210: contig of 748 bp in length
33211 33310: gap of 100 bp
33311 34076: contig of 766 bp in length
34077 34176: gap of 100 bp
34177 34935: contig of 759 bp in length
34936 35035: gap of 100 bp
35036 35756: contig of 721 bp in length
35757 35856: gap of 100 bp
35857 36585: contig of 729 bp in length
36586 36685: gap of 100 bp
36686 37431: contig of 746 bp in length
37432 37531: gap of 100 bp
37532 38275: contig of 744 bp in length
38276 38375: gap of 100 bp

```

* 38376 39094: contig of 719 bp in length
* 39095 39194: gap of 100 bp
* 39195 39960: contig of 766 bp in length
* 39961 40060: gap of 100 bp
* 40061 40779: contig of 719 bp in length
* 40780 40879: gap of 100 bp
* 40880 41612: contig of 733 bp in length
* 41613 41712: gap of 100 bp
* 41713 42441: contig of 729 bp in length
* 42442 42541: gap of 100 bp
* 42542 43275: contig of 734 bp in length
* 43276 43375: gap of 100 bp
* 43376 44114: contig of 739 bp in length
* 44115 44214: gap of 100 bp
* 44215 44965: contig of 751 bp in length
* 44966 45065: gap of 100 bp
* 45066 45789: contig of 724 bp in length
* 45790 45889: gap of 100 bp
* 45890 46633: contig of 744 bp in length
* 46634 46733: gap of 100 bp
* 46734 47462: contig of 729 bp in length
* 47463 47562: gap of 100 bp
* 47563 48301: contig of 739 bp in length
* 48302 48401: gap of 100 bp
* 48402 49125: contig of 724 bp in length
* 49126 49225: gap of 100 bp
* 49226 49962: contig of 737 bp in length
* 49963 50062: gap of 100 bp
* 50063 50798: contig of 736 bp in length
* 50799 50898: gap of 100 bp
* 50899 51635: contig of 737 bp in length
* 51636 51735: gap of 100 bp
* 51736 52464: contig of 729 bp in length
* 52465 52564: gap of 100 bp
* 52565 53316: contig of 753 bp in length
* 53317 53416: gap of 100 bp
* 53417 54176: contig of 760 bp in length
* 54177 54276: gap of 100 bp
* 54277 55004: contig of 728 bp in length
* 55005 55104: gap of 100 bp
* 55105 55867: contig of 763 bp in length
* 55868 55967: gap of 100 bp
* 55968 56676: contig of 709 bp in length
* 56677 56776: gap of 100 bp
* 56777 57511: contig of 735 bp in length
* 57512 57611: gap of 100 bp

```

Query Match 1.2% Score 20: DB 2: Length 74274:

Best Local Similarity 100.0%: Pred. No. 36: Mismatches 0: Indels 0: Gaps 0:

Oy 187 ATACAAAGACACTCTTATT 206
 Db 43542 ATACAAAGACACTCTTATT 43561

RESULT 40
 AC025723 77422 bp DNA linear INV 29-MAY-2002
 LOCUS Caenorhabditis elegans cosmid Y54F10AM, complete sequence.
 AC025723
 AC025723.2 GI:13559726
 VERSION HTG.
 KEYWORDS Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 77422)
 Waterston, R.
 REFERENCE Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 JOURNAL Science 282 (5396), 2012-2018 (1998)
 MEDLINE 99069613

PUBMED 9851916
 REFERENCE 2 (bases 1 to 77422)
 AUTHORS Ryan, E., Courtney, L. and Yoakum, M.
 TITLE The sequence of C. elegans cosmid Y54F10AM
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 77422)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAR-2000) Genome Sequencing Center, Washington
 University, School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 77422)
 REFERENCE Waterston, R.
 AUTHORS Direct Submission
 TITLE Submitted (06-APR-2001) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 JOURNAL 5 (bases 1 to 77422)
 REFERENCE Waterston, R.
 AUTHORS Direct Submission
 TITLE Submitted (05-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 JOURNAL 6 (bases 1 to 77422)
 REFERENCE Waterston, R.
 AUTHORS Direct Submission
 TITLE Submitted (29-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 JOURNAL On Apr 6, 2001 this sequence version replaced gi:7230347.
 COMMENT Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1HQ, England
 email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats: all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=Y54F10AM;class=Sequenc>
 e

NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y54F10AL, 6000 bp overlap: the 3' cosmid is
 Y54F10AR, 200 bp overlap.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 GeneFinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yoji Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFeome cloning project (<http://worfdb.dicf.harvard.edu/>),
 similarity to other proteins from Blastx analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10-1115-1125, 2000), individual C. elegans GenBank submissions,
 and personal communications with C. elegans researchers. tRNAs
 are predicted using the program tRNAscan-SE (Lowe, T.M. and

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 20:35:19 ; Search time 1670 Seconds
(without alignments)
15875.456 Million cell updates/sec

Title: US-09-869-433-1
Perfect score: 1637
Sequence: 1 gaataaataactacacgaa.....gttgatccctacgtcttt 1637

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1.5	755	17	BH837547	BH837547 LMCOR05001
2	1.4	483	17	BH773218	BH773218 uzmbo002fo
3	1.3	770	17	AQ893684	AQ893684 HS_3148_A
4	1.3	292	9	AA639627	AA639627 ng83d01.s
5	1.3	330	17	AQ013184	AQ013184 CIT-HSP-2
6	1.3	346	9	A1002102	A1002102 ct38d04.s

7	21	1.3	346	10	AM090585
8	21	1.3	420	14	B0012918
9	21	1.3	509	9	AA644559
10	21	1.3	597	17	A2865523
11	21	1.3	951	21	BH164112
12	21	1.3	1017	14	B0048904
13	21	1.2	119	14	AQ115988
14	20	1.2	254	14	B0610731
15	20	1.2	289	10	BA36799
16	20	1.2	297	10	AM374096
17	20	1.2	303	10	BB216685
18	20	1.2	312	10	BB381288
19	20	1.2	328	9	AA712844
20	20	1.2	334	17	AQ697744
21	20	1.2	337	17	AZ016529
22	20	1.2	339	10	BE432018
23	20	1.2	342	12	AQ685833
24	20	1.2	352	17	BG628369
25	20	1.2	356	12	BG626629
26	20	1.2	368	12	BG628526
27	20	1.2	388	10	AM011059
28	20	1.2	439	17	A2630188
29	20	1.2	439	10	AM218715
30	20	1.2	440	10	AM218716
31	20	1.2	458	10	AM933403
32	20	1.2	472	17	AQ433974
33	20	1.2	496	12	BF098187
34	20	1.2	498	10	AM651372
35	20	1.2	500	13	BM404442
36	20	1.2	514	10	AM221154
37	20	1.2	546	10	BB284364
38	20	1.2	548	17	AQ121410
39	20	1.2	558	14	BQ119131
40	20	1.2	559	10	BE459902
41	20	1.2	566	10	AM907382
42	20	1.2	569	9	A1993855
43	20	1.2	575	14	BQ116413
44	20	1.2	576	13	BM112431
45	20	1.2	579	12	BE919839
46	20	1.2	580	10	BE41982
47	20	1.2	582	10	BE460070
48	20	1.2	585	10	AM035953
49	20	1.2	589	10	BE451607
50	20	1.2	594	17	A2798455
51	20	1.2	595	10	AM930772
52	20	1.2	596	10	AV751415
53	20	1.2	596	10	AM221000
54	20	1.2	597	10	AM932225
55	20	1.2	604	17	A2821248
56	20	1.2	619	9	A1771102
57	20	1.2	619	12	BG644061
58	20	1.2	623	14	BQ046082
59	20	1.2	629	13	BM113917
60	20	1.2	633	12	BG589972
61	20	1.2	648	17	A2625761
62	20	1.2	657	10	AM773675
63	20	1.2	661	10	BB150648
64	20	1.2	666	17	AG053391
65	20	1.2	671	12	BG594974
66	20	1.2	676	10	BE461454
67	20	1.2	685	10	AM040455
68	20	1.2	685	17	A2338986
69	20	1.2	690	13	BI922435
70	20	1.2	695	13	BI406811
71	20	1.2	713	13	BI923527
72	20	1.2	719	13	BM410984
73	20	1.2	720	12	BG890033
74	20	1.2	725	12	BG598397
75	20	1.2	725	17	AQ247939
76	20	1.2	732	12	BG125387
77	20	1.2	771	12	BF573280
78	20	1.2	844	14	BQ510145
79	20	1.2	850	17	A2540960

```

80 20 1.2 888 17 A2546199
81 20 1.2 902 17 BH149116
82 20 1.2 1068 17 BH663206
83 19 1.2 195 14 BQ234495
84 19 1.2 283 12 BF516008
85 19 1.2 296 10 AM435908
86 19 1.2 304 10 A2790186
87 19 1.2 313 10 BB208477
88 19 1.2 353 17 AQ237639
89 19 1.2 359 17 AQ302835
90 19 1.2 360 9 AA591598
91 19 1.2 383 17 AQ145028
92 19 1.2 399 9 AA516865
93 19 1.2 401 10 AM815196
94 19 1.2 406 17 AQ047250
95 19 1.2 411 10 AM390055
96 19 1.2 414 10 BE090450
97 19 1.2 417 9 AA767956
98 19 1.2 418 9 BB686390
99 19 1.2 420 9 AA462906
C 100 19 1.2 426 10 AM208986

```

ALIGNMENTS

```

RESULT 1
BH837547/c 755 bp DNA linear GSS 28-MAY-2002
LOCUS LMKR050015H09f Zea mays L. Zea mays genomic clone LMKR050015H09f,
DEFINITION
DNA sequence.
ACCESSION BH837547
VERSION BH837547.1 GI:21235425
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 755)
Kim.S.W., Yu.Y., Lee.M.C., Main.D. and Wang.R.A.
Methyl-filtration genomic sequence from maize
JOURNAL Unpublished (2002)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total High Quality bases = 520
Seq primer: TAATPACGACCTCATPACG
Class: shotgun
High quality sequence start: 12
High quality sequence stop: 714.
Location/Qualifiers
1..755

```

```

FEATURES
source
1..755

```

```

FEATURES
source
1..755
/organism="Zea mays"
/strain="B73"
/db_xref="taxon:4577"
/clone="LMKR050015H09f"
/tissue_type="Zea mays L."
/label="leaf"
/lab_host="DH10B"
/note="Vector: pGEM-T easy; Site_1: Mcr BC;
Methyl-filtration library; Nuclei DNA was completely
digested with Mcr BC, size fractionated and transformed
to E.Coli DH10B."

```

```

BASE COUNT 242 a 176 c 152 g 179 t 6 others
ORIGIN
Query Match 1.5%; Score 25; DB 17; Length 755;
Best Local Similarity 100.0%; Pred. No. 0.36;

```

```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 20 ATGAGAAATTAAGATTTCAGAG 44
|||||
Db 299 ATGAGAAATTAAGATTTCAGAG 275

```

```

RESULT 2
BH773218 483 bp DNA linear GSS 28-MAR-2002
LOCUS uzmb002f003c09 5', DNA sequence.
DEFINITION
uzmb002f003c09f uzmb unfiltered library Zea mays genomic clone
GSS.
ACCESSION BH773218
VERSION BH773218.1 GI:19775208
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 483)
Budiman,M.A., Freese,R.G., Bedell,J.A., Nunberg,A.N. and Lakey,N.D.
Genethresher methylation filtered genomic sequences from maize
JOURNAL Unpublished (2002)
COMMENT Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: uzmb002f003 row: c column: 09
Seq primer: M13 forward
Class: shotgun
High quality sequence stop: 483.
Location/Qualifiers
1..483

```

```

FEATURES
source
1..483

```

```

FEATURES
source
1..483
/organism="Zea mays"
/cultivar="MO17"
/db_xref="taxon:4577"
/clone="uzmb002f003c09"
/note="Organ: leaf; Vector: pGEM-T(-); Site_1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to
5 kb fraction, ligated into HincII-digested pGEM-T(-)
vector and electroporated into E. coli cells."
BASE COUNT 128 a 98 c 93 g 164 t
ORIGIN
Query Match 1.4%; Score 23; DB 17; Length 483;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 20 ATGAGAAATTAAGATTTCAGAG 42
|||||
Db 428 ATGAGAAATTAAGATTTCAGAG 450

```

```

RESULT 3
AO893684/c 770 bp DNA linear GSS 10-NOV-1999
LOCUS HS_3148_A2.C11_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3148 COL=22 Row=E, DNA sequence.
ACCESSION AO893684
VERSION AO893684.1 GI:6349874
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 770)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

```

```

TITLE
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL
99380589
MEDLINE
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
BAC end Web Server: http://www.husc.washington.edu
Plate: 3148 row: E column: 22
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 770.
FEATURES
1. 770
Location/Qualifiers
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3148 Col=22 Row=E"
/clone.lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT
245 a 144 c 108 g 273 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AAGAAAAACCTTTGGAAAAATT 85
|||||
Db 702 AAGAAAAACCTTTGGAAAAATT 681

RESULT 4
AA639627 292 bp mRNA linear EST 23-OCT-1997
ng83d01.sl NCI-CGAP_Co9 Homo sapiens cDNA IMAGE:1158913 3',
DEFINITION
mRNA sequence.
ACCESSION
AA639627
VERSION
AA639627.1 GI:2563406
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 292)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/BLNT at:
www.bio.lnlnl.gov/bdrp/image/image.html
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 282.
FEATURES
1. 292
Location/Qualifiers
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1158913"

```

```

/clone.lib="NCI CGAP Co9"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(drf) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo (Soares4
).
BASE COUNT
109 a 63 c 37 g 83 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 33;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 TAGATCCTGGGGAATTTCTT 720
|||||
Db 289 TAGATCCTGGGGAATTTCTT 269

RESULT 5
AA013184 330 bp DNA linear GSS 06-JUN-1998
CIT-HSP-2298E18.TR CIT-HSP Homo sapiens genomic clone 2298E18, DNA
LOCUS
sequence.
DEFINITION
AA013184
ACCESSION
AA013184
VERSION
AA013184.1 GI:3185749
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 330)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other_GSSs: CIT-HSP-2298E18.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
1. 330
Location/Qualifiers
source
/organism="Homo sapiens"
/db_xref="GDB:7153678"
/db_xref="taxon:9606"
/clone="2298E18"
/clone.lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT
112 a 58 c 67 g 93 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


Clone distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@iowa.edu
 Seq primer: M13 FORWARD
 POLYA=Yes

FEATURES

Source

1. 420
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-1-BC1p-ava-e-07-0-UI"
 /clone_lib="NCLCGAP.pl3"
 /tissue_type="Placenta"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B (Life Technologies)"
 /note="Organ: Placenta; Vector: pT73-Pac (pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; NCLCGAP.pl3 is a subcloned cDNA library constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are GA, AGGA. For additional information, contact: Bento Soares, bento-soares@iowa.edu
 TAG_Lib=UI-1-BC1p
 TAG_Tissue=placenta human 8 week
 TAG_SEQ=GA"

BASE COUNT 143 a 83 c 46 g 148 t

ORIGIN

Query Match 1.3%; Score 21; DB 14; Length 420;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 700 TAGATCCTGGGGAATTCTT 720
 ||||||||||||||||
 Db 308 TAGATCCTGGGGAATTCTT 288

RESULT 9
 AA644559/c 509 bp mRNA linear EST 06-MAR-1998
 LOCUS ab63908.s1 StrataGene lung carcinoma 937218 Homo sapiens cDNA clone
 DEFINITION IMAGE:845534 3', mRNA sequence.
 ACCESSION AA644559
 VERSION AA644559.1 GI:2569777
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 509)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Merra,M., Martin,J., Moore,B., Schellberg,K., Stepcoe,M., Tan,F., Theisling,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 TITLES WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

EMAIL: est@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1206 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 463.
 Location/Qualifiers

FEATURES

source

1. 509
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:845534"
 /clone_lib="StrataGene lung carcinoma 937218"
 /tissue_type="lung Carcinoma"
 /cell_line="NCI-H69"
 /dev_stage="cell line NCI-H69"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: lung; Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Small cell carcinoma cell line NCI-H69. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAGC 3' -3' adaptor sequence: 5' CTCAGGTTTTTTTTTTTTTTT 3' "

BASE COUNT 173 a 107 c 67 g 162 t

ORIGIN

Query Match 1.3%; Score 21; DB 9; Length 509;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 700 TAGATCCTGGGGAATTCTT 720
 ||||||||||||||||
 Db 292 TAGATCCTGGGGAATTCTT 272

RESULT 10
 A2865523 597 bp DNA linear GSS 21-FEB-2001
 LOCUS 2M0175K12P Mouse 10kb plasmid U06C1M library Mus musculus genomic
 DEFINITION clone U06C2M0175K12 R, DNA sequence.
 ACCESSION A2865523
 VERSION A2865523.1 GI:13065920
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 597)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
 TITLES Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

EMAIL: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0175 row: K column: 12
 Seq primer: CACACAGGAACACCTAGACC
 Class: plasmid ends
 High quality sequence stop: 597.
 Location/Qualifiers

FEATURES

Source

1. 597
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U06C2M0175K12"
 /clone_lib="Mouse 10kb plasmid U06C1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42mv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (911473211419b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 194 a 148 c 89 g 166 t

Query Match 1.3%; Score 21; DB 17; Length 597;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 873 AAGATGAGCTGCTCTATCTT 893
|||||
DB 184 AAGATGAGCTGCTCTATCTT 204

RESULT 11

BH164112 951 bp DNA linear GSS 24-SEP-2001
LOCUS ENTB33Tf Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.

ACCESSION BH164112
VERSION BH164112.1 GI:15737550
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 951)
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)

JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 15
High quality sequence stop: 754.
Location/Qualifiers

FEATURES
source 1..951
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_id="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O., and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 452 a 86 c 132 g 281 t

OY 18 GAATGAAATTAAGTATTT 38
|||||
DB 717 GAATGAAATTAAGTATTT 737

RESULT 12 1017 bp mRNA linear EST 29-MAR-2002
LOCUS BQ048904
DEFINITION AGENCOURT_6832569 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5789594
5', mRNA sequence.
ACCESSION BQ048904
VERSION BQ048904.1 GI:19808244
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1017)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM12883 row: 9 column: 03
High quality sequence stop: 699.
Location/Qualifiers

FEATURES
source 1..1017
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:5789594"
/clone_id="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."

BASE COUNT 339 a 195 c 219 g 264 t

Query Match 1.3%; Score 21; DB 14; Length 1017;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 AGAAAAACCTTTGGAAATT 85
|||||
DB 985 AGAAAAACCTTTGGAAATT 1005

RESULT 13 119 bp DNA linear GSS 20-APR-1999
LOCUS A0115988
DEFINITION RC111-57F1.TK.1 RPCR-11 Homo sapiens genomic clone RPCR-11-57F1, DNA sequence.
ACCESSION A0115988
VERSION A0115988.1 GI:3492109

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Liner,K., Golden,K., Berry,K., Granger,D., Suh,E., Mble,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other-GSSs: RPC11-57P1.TJB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@ejong.med.buffalo.edu). Clones may be purchased from BACPac Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Class: BAC ends.

FEATURES
source
1..119
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7521624"
/db_xref="taxon:9606"
/clone="RPC1-11-57F1"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC library"

BASE COUNT
ORIGIN
40 a 28 c 19 g 32 t

Query Match 1.2%; Score 20; DB 17; Length 119;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1291 CTCCTTTGACTCACTAA 1310
Db 16 CTCCTTTGACTCACTAA 35

RESULT 14
BO610731/c 254 bp mRNA linear EST 26-JUN-2002
DEFINITION B0610731
saps0605.y1 Gm-cl087 glycine max cDNA clone SOYBEAN CLONE ID:
Gm-cl087-3873 5', mRNA sequence.
ACCESSION B0610731
VERSION B0610731
KEYWORDS B0610731.1 GI:21600400
SOURCE EST.
ORGANISM soybean.
glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 254)
AUTHORS Shoemaker,R., Keim,P., Voklin,L., Erpeiding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,D., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theissen,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Hunttsville, AL 35801 For further information call: (800)-533-4363 or contact: cc@resgen.com web site: www.resgen.com

FEATURES
source
1..254
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl087-3873"
/clone_lib="Gm-cl087"
/tissue_type="Roots"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The mRNA was prepared using polyattract mRNA system from PROMEGA. The cDNA was prepared using the STRATAGENE kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site
(5'GACGACAGAGAGAGACAGACTGCTCGAGTTTGTCTTTTGTCTTTT).
EcoRI adaptors (5'OH-AATTGGCGACGAG and 3'GCCGCTCTCP) were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction sites of the pluescript vector. The vector was previously dephosphorylated. The ligated cDNA fragments were transformed into DH10B by electroporation. Library is in LB medium with ampicillin and glycerol 8%. Average insert size: 800 bp."

BASE COUNT
ORIGIN
81 a 42 c 47 g 84 t

Query Match 1.2%; Score 20; DB 14; Length 254;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 115 ACGAGCTAAGAGAAAGTTCTG 134
Db 186 ACGAGCTAAGAGAAAGTTCTG 167

RESULT 15
BB436799 289 bp mRNA linear EST 02-AUG-2000
DEFINITION BB436799
musculus cDNA clone C82020F15 3', mRNA sequence.
ACCESSION BB436799
VERSION BB436799
KEYWORDS BB436799.1 GI:9276526
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 289)
AUTHORS Kono,H., Alzawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,T., Ishikawa,U., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,T., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomiなが,N., Toya,T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamataka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Kono,H., et al.)

JOURNAL
COMMENT
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermocatalytic activation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1. .289
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G820020F15"
/clone_lib="RIKEN full-length enriched, adult pancreas
islet cells"
/tissue_type="pancreas"
/cell_type="islet cells"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI. Site_2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
GAGAGAGAGAGATCCAGAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATCTCCAGTTAATTAAATTATATCCCCCCCC 3']. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KS(+) after bulk excision from lambda FLC
I.-Islet cells were provided by Hiroo Iwata, Institute for
Frontier Medical Sciences, Kyoto University, Sakyo-ku,
Kyoto, 606-8507 Japan, whose assistance we gratefully
acknowledge."

BASE COUNT 82 a 74 c 44 g 89 t
ORIGIN

Query Match 1.2%: Score 20; DB 10; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TTGTGGAGCTCTGCTGCT 224
|||||
Db 73 TTGTGGAGCTCTGCTGCT 92

RESULT 16
LOCUS AM374096 297 bp MRNA linear EST 04-FEB-2000
DEFINITION PM4-BT0548-L71299-001-g12 BT0548 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM374096
VERSION AM374096.1 GI:6878750

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
1 (bases 1 to 297)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM4&t2=PM4-BT0548-
171299-001-g12&t3=1999-12-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 20
High quality sequence stop: 172.
FEATURES
source
Location/Qualifiers
1. .297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0548"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 89 a 66 c 38 g 104 t
ORIGIN

Query Match 1.2%: Score 20; DB 10; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 TGCATATCTCAATATGAT 1010
|||||
Db 175 TGCATATCTCAATATGAT 194

RESULT 17
LOCUS BB216685/c 303 bp MRNA linear EST 30-JUN-2000
DEFINITION BB216685 RIKEN full-length enriched, adult male aorta and vein Mus
musculus cDNA clone A530037K10 3', mRNA sequence.
ACCESSION BB216685
VERSION BB216685.1 GI:8881638
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 303)
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toyu,
T., Tsunoda, Y., Wataniki, A., Watanabe, S., Yamamura, T., Yamana, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino

TITLE
JOURNAL
COMMENT

., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermocatalysis of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

SOURCE

1. 303
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A530037K10"
/clone_lib="RIKEN full-length enriched, adult male aorta
and vein"
/sex="male"
/tissue_type="aorta and vein"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 78 a 70 c 46 g 109 t
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 303;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1336 AAGAGCAAAAGTCAAGGT 1355
Db 120 AAGAGCAAAAGTCAAGGT 101

RESULT 18
LOCUS BB381288 312 bp mRNA linear EST 13-JUL-2000
DEFINITION BB381288 RIKEN full-length enriched, 0 day neonate cerebellum Mus
musculus cDNA clone C230012B13 3', mRNA sequence.
ACCESSION BB381288

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

BB381288.1 GI:9101012
EST
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 312)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Atake, T., Carninci,
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomihaga, N., Toya,
T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermocatalysis of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

SOURCE

1. 312
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230012B13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 479.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 103 a 45 c 53 g 111 t

```

ORIGIN
Query Match      1.2%; Score 20; DB 10; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 TAATGAATTTACAAAATCC 585
      |||||||
Db 171 TAATGAATTTACAAAATCC 190

RESULT 19
AA712844 328 bp mRNA linear EST 24-DEC-1997
DEFINITION 32403 Lambda-PRL2 Arabidopsis thaliana cDNA clone 201N10T7, mRNA
sequence.
ACCESSION AA712844
VERSION AA712844.1 GI:2722761
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 328)
AUTHORS Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
,E. and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant. Physiol. 106, 1241-1255 (1994)
JOURNAL MEDLINE
COMMENT Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel.: 517-353-0854
Fax: 517-353-9168
Email: 22313cnc@lhm.cl.msu.edu
Seq primer: T7 dye primer.

FEATURES
Source Location/Qualifiers
1..328
/organism="Arabidopsis thaliana"
/strain="var Columbia"
/db_xref="taxon:3702"
/clone="201N10T7"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."
BASE COUNT 80 a 63 c 81 g 91 t 13 others
ORIGIN
Query Match      1.2%; Score 20; DB 9; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1308 AAAGAAATGCGCTATATCCC 1327
      |||||||
Db 80 AAAGAAATGCGCTATATCCC 99

RESULT 20
AO697744 334 bp DNA linear GSS 06-JUL-1999
LOCUS AO697744

```

```

DEFINITION HS_5553_A2-D03.SP6E RPCI-11 Human Male BAC library Homo sapiens
genomic clone plate-1129 Col=6 Row=G, DNA sequence.
ACCESSION AO697744
VERSION AO697744.1 GI:5387992
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL MEDLINE
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (Info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1129 row: G column: 6
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 334.

FEATURES
Source Location/Qualifiers
1..334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=1129 Col=6 Row=G"
/clone_lib="RPCI-11 Human Male BAC library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT 127 a 53 c 63 g 91 t
ORIGIN
Query Match      1.2%; Score 20; DB 17; Length 334;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AAATPAAAACTATCAGAT 21
      |||||||
Db 115 AAATPAAAACTATCAGAT 134

RESULT 21
AZ016529/c 337 bp DNA linear GSS 25-FEB-2000
LOCUS AZ016529
DEFINITION RPCI-23-368M9.TJ RPCI-23 Mus musculus genomic clone RPCI-23-368M9,
DNA sequence.
ACCESSION AZ016529
VERSION AZ016529.1 GI:7091913
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Zhao,S., Nierman,W., Felblyum,T., Malek,J., Shatsman,S., Akliuret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.

```

TITLE
JOURNAL
COMMENT

and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_CSSS: RPCI-23-368M9.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@ligr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.ligr.org/rdb/bac_ends/mouse/bac_end_intro.html
Plate: 368 row: M column: 9
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..337
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-368M9"
/clone_lib="RPCI-23"
/sex="female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6j mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT
135 a 59 c 41 g 102 t
ORIGIN

Query Match
Best Local Similarity 1.2%; Score 20; DB 17; Length 337;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 ATTACTTATGCTTATTTA 295
|||||
DB 323 ATTACTTATGCTTATTTA 304

RESULT 22
BE432018/c 339 bp mRNA linear EST 18-MAY-2001
DEFINITION ES1398547 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
LOCUS BE432018
ACCESSION BE432018
VERSION BE432018.2 GI:11387655
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 339)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
On Jul 24, 2000 this sequence version replaced gi:9429861.
Contact: CUGI
Clemson University Genomics Institute
Clemson University

TITLE
JOURNAL
COMMENT

100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
Location/Qualifiers
1..339
/organism="Lycopersicon esculentum"
/cultivar="A496"
/db_xref="taxon:4081"
/clone="CLEG4L20"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmClnadap; Site_1: EcoRI; Site_2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopen accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."
BASE COUNT
101 a 71 c 50 g 117 t
ORIGIN

Query Match
Best Local Similarity 1.2%; Score 20; DB 10; Length 339;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1571 AGAAGATTGAGTCCTGCTT 1590
|||||
DB 65 AGAAGATTGAGTCCTGCTT 46

RESULT 23
A0685833 342 bp DNA linear GSS 28-JUN-1999
DEFINITION HS.5549.AL.G08-SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-1125 Col-15 Row-M, DNA sequence.
LOCUS A0685833
ACCESSION A0685833
VERSION A0685833.1 GI:5261904
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 342)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resea ch Genetics (info@resgen.com). BAC end web Server: http://www.htsc.washington.edu
Plate: 1125 row: M column: 15
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 342.
Location/Qualifiers
1..342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1125 Col-15 Row-M"
/clone_lib="RPCI-11 Human Male BAC Library"

/sex="male"
/note="vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"

BASE COUNT 118 a 55 c 58 g 111 t

Query Match 1.2%; Score 20; DB 17; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 AATGTTCCATATGTTCTTCT 156
|||||
DB 137 AATGTTCCATATGTTCTTCT 156

RESULT 24 352 bp mRNA linear EST 19-APR-2001
BG628369
LOCUS CC-estf1cEL21p10a1 Tomato flower library from a mixture of
developmental stages Lycopersicon esculentum cDNA clone
CC-estf1cEL21p10a1, mRNA sequence.

ACCESSION BG628369 GI:13679842
VERSION EST.
KEYWORDS tomato.
SOURCE Lycopersicon esculentum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 352)
AUTHORS van der Hoeven, R.S. and Tanksley, S.D.
TITLE ESTs from a tomato flower library
JOURNAL Unpublished (2001)
COMMENT Contact: Rutgers S. van der Hoeven
Cornell University
252 Emerson Hall, Ithaca, NY 14850, USA
Tel: 607 255 7866
Fax: 607 255 6683
Email: rv19@cornell.edu
3 prime sequence.

FEATURES Location/Qualifiers

1..352
/organism="Lycopersicon esculentum"
/cultivar="E6203"
/db_xref="taxon:4081"
/clone="CC-estf1cEL21p10a1"
/clone_1lb="Tomato flower library from a mixture of
developmental stages"
/tissue_type="developing flower buds and open flowers"
/dev_stage="4-8 week old plants"
/lab_host="X10LR"
/note="Vector: pBK_CMV; Site_1: EcoRI; Site_2: XhoI;
Flowers and flower buds were collected from greenhouse
grown plants and used for library construction (cEL)."

BASE COUNT 102 a 55 c 68 g 127 t

Query Match 1.2%; Score 20; DB 12; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAAGATTCAGCTCCTGCTT 1590
|||||
DB 252 AGAAGATTCAGCTCCTGCTT 271

RESULT 25 356 bp mRNA linear EST 19-APR-2001
BG626629
LOCUS

DEFINITION CC-estf1cEL21p10a1 Tomato flower library from a mixture of
developmental stages Lycopersicon esculentum cDNA clone
CC-estf1cEL21p10a1, mRNA sequence.

ACCESSION BG626629 GI:13678102
VERSION EST.
KEYWORDS tomato.
SOURCE Lycopersicon esculentum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 356)
AUTHORS van der Hoeven, R.S. and Tanksley, S.D.
TITLE ESTs from a tomato flower library
JOURNAL Unpublished (2001)
COMMENT Contact: Rutgers S. van der Hoeven
Cornell University
252 Emerson Hall, Ithaca, NY 14850, USA
Tel: 607 255 7866
Fax: 607 255 6683
Email: rv19@cornell.edu
3 prime sequence.

FEATURES Location/Qualifiers

1..356
/organism="Lycopersicon esculentum"
/cultivar="E6203"
/db_xref="taxon:4081"
/clone="CC-estf1cEL21p10a1"
/clone_1lb="Tomato flower library from a mixture of
developmental stages"
/tissue_type="developing flower buds and open flowers"
/dev_stage="4-8 week old plants"
/lab_host="X10LR"
/note="Vector: pBK_CMV; Site_1: EcoRI; Site_2: XhoI;
Flowers and flower buds were collected from greenhouse
grown plants and used for library construction (cEL)."

BASE COUNT 103 a 54 c 74 g 125 t

Query Match 1.2%; Score 20; DB 12; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAAGATTCAGCTCCTGCTT 1590
|||||
DB 261 AGAAGATTCAGCTCCTGCTT 280

RESULT 26 368 bp mRNA linear EST 19-APR-2001
BG628526
LOCUS CC-estf1cEL22116d1 Tomato flower library from a mixture of
developmental stages Lycopersicon esculentum cDNA clone
CC-estf1cEL22116d1, mRNA sequence.

ACCESSION BG628526 GI:13679999
VERSION EST.
KEYWORDS tomato.
SOURCE Lycopersicon esculentum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 368)
AUTHORS van der Hoeven, R.S. and Tanksley, S.D.
TITLE ESTs from a tomato flower library
JOURNAL Unpublished (2001)
COMMENT Contact: Rutgers S. van der Hoeven
Cornell University
252 Emerson Hall, Ithaca, NY 14850, USA
Tel: 607 255 7866
Fax: 607 255 6683

Email: rv19@cornell.edu
3 prime sequence.

FEATURES
source
Location/Qualifiers
1..368

/organism="Lycopersicon esculentum"
/cultivar="E6203"
/db_xref="taxon:4081"
/clone="cc-esf1cLEL22116d1"
/clone_1lb="tomato flower library from a mixture of developmental stages"
/tissue_type="developing flower buds and open flowers"
/dev_stage="4-8 week old plants"
/lab_host="XL0LR"
/note="Vector: PBK CMV; Site 1: EcoRI; Site 2: XhoI; flowers and flower buds were collected from greenhouse grown plants and used for library construction (cLEL)."

BASE COUNT
ORIGIN
106 a 56 c 75 g 131 t

Query Match
Best Local Similarity 100.0%; Pred. No. 1.le+02; Length 368;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1571 AGAAGATTCAGCTCTGCTT 1590
Db 262 AGAAGATTCAGCTCTGCTT 281

RESULT 27
LOCUS
DEFINITION
AM011059 388 bp mRNA linear EST 10-SEP-1999
S716C01 Pine Triplex shoot tip library Pinus taeda cDNA clone
AM011059 mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AM011059
AM011059.1 GI:5859837
EST.
loblolly pine.
Pinus taeda

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus. 1 (bases 1 to 388)
Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
The Pine Gene Discovery Project
Unpublished (1999)
Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhetten@ncsu.edu

FEATURES
source
Seq primer: 5' Lambda Triplex2 Sequencing Primer.
Location/Qualifiers
1..388
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="S716C01"
/clone_1lb="Pine Triplex shoot tip library"
/lab_host="E. coli BM25.8"
/note="Organ: shoot tips; Vector: Lambda Triplex; Site 1: SfiI (A); Site 2: SfiI (B); Shoot tips (approx. 2 cm from apex) were collected during the spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT
ORIGIN
101 a 62 c 92 g 126 t 7 others

Query Match
Best Local Similarity 100.0%; Pred. No. 1.le+02; Length 388;
1.2%; Score 20; DB 10; Length 388;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1494 ATTGCGTTCGTTCTGCAC 1513
Db 123 ATTGGTTGTTCTGCAC 142

RESULT 28
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
A2630188 439 bp DNA linear GSS 13-DEC-2000
IM0483D12R Mouse 10kb plasmid U06C1M library Mus musculus genomic clone U06C1M0483D12 R, DNA sequence.
A2630188
A2630188.1 GI:11752378
GSS.
house mouse.
Mus musculus
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0483 row: D column: 12
Seq primer: CACACAGAAACACCTATGACC
Class: plasmid ends
High quality sequence stop: 439.
Location/Qualifiers
1..439
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0483D12"
/clone_1lb="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD2 (g114732114[gb]AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

FEATURES
source

BASE COUNT
ORIGIN
140 a 69 c 53 g 177 t

Query Match
Best Local Similarity 100.0%; Pred. No. 1.le+02; Length 439;
1.2%; Score 20; DB 17; Length 439;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 851 TGCTAACTAAATGATA 870
|||||
Db 13 TGCTAACTAAATGATA 32

RESULT 29
AM218715/c 440 bp mRNA linear EST 18-MAY-2001
LOCUS EST301195 tomato root during/after fruit set, Cornell University
DEFINITION Lycopersicon esculentum cDNA clone cLEXIA3, mRNA sequence.
ACCESSION AM218715
VERSION AM218715.1 GI:6529589
KEYWORDS EST.
SOURCE Lycopersicon esculentum
ORGANISM tomato.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 440)
van der Hoeven,R.S., Garvin,D., Matera,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Romning,C.M.,
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato root tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1..440
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEXIA3"
/clone_lib="tomato root during/after fruit set, Cornell
University"
/tissue_type="root"
/dev_stage="plants during and after fruit-set"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin
(USDA-ARS, Ithaca, NY 14850)."
BASE COUNT 133 a 92 c 71 g 144 t
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1571 AGAAGATTACGCTCCTT 1590
|||||
Db 166 AGAAGATTACGCTCCTT 147

RESULT 30
AM218716 440 bp mRNA linear EST 18-MAY-2001
LOCUS EST301196 tomato root during/after fruit set, Cornell University
DEFINITION Lycopersicon esculentum cDNA clone cLEXIA3, mRNA sequence.
ACCESSION AM218716
VERSION AM218716.1 GI:6529590
KEYWORDS EST.
SOURCE Lycopersicon esculentum
ORGANISM tomato.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 440)

AUTHORS van der Hoeven,R.S., Garvin,D., Matera,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Romning,C.M.,
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato root tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
3 prime sequence.
Location/Qualifiers
1..440
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEXIA3"
/clone_lib="tomato root during/after fruit set, Cornell
University"
/tissue_type="root"
/dev_stage="plants during and after fruit-set"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin
(USDA-ARS, Ithaca, NY 14850)."
BASE COUNT 144 a 71 c 92 g 133 t
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1571 AGAAGATTACGCTCCTT 1590
|||||
Db 275 AGAAGATTACGCTCCTT 294

RESULT 31
AM933403/c 458 bp mRNA linear EST 18-MAY-2001
LOCUS EST359342 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF54010 5', mRNA sequence.
ACCESSION AM933403
VERSION AM933403.1 GI:8108900
KEYWORDS EST.
SOURCE Lycopersicon esculentum
ORGANISM tomato.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 458)
Alcala,J., Vrebalov,J., White,R., Matera,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Romning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1..458
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF54010"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XhoI: CLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'Immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 141 a 103 c 99 g 115 t
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAAGATTCAGCTCCTGCTT 1590
|||||
Db 381 AGAAGATTCAGCTCCTGCTT 362

RESULT 32
AQ433974

DEFINITION HS.5088_A1_D04_SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=664 Col=7 Row=G, DNA sequence.

ACCESSION AQ433974
VERSION AQ433974.1 GI:4544309

KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 472)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 664 Row: G Column: 7
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 472.

FEATURES
Source Location/Qualifiers
1..472

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=664 Col=7 Row=G"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

BASE COUNT 166 a 75 c 91 g 139 t 1 others
ORIGIN

Query Match 1.2%; Score 20; DB 17; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AAATAAAACTATCAGAT 21
|||||
Db 120 AAATAAAACTATCAGAT 139

RESULT 33
BF098187/c

DEFINITION BF098187 496 bp mRNA linear EST 18-MAY-2001
EST428796 tomato nutrient deficient roots Lycopersicon esculentum
cDNA clone cLEM26C18 5' sequence, mRNA sequence.

ACCESSION BF098187
VERSION BF098187.1 GI:10903985

KEYWORDS EST.
SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 496)
AUTHORS van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T.S., Rohning,C.M., Craven,M.B., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato nutrient deficient roots
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.

TITLE Generation of ESTs from tomato nutrient deficient roots
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
Source Location/Qualifiers
1..496

/organism="Lycopersicon esculentum"
/cultivar="TA492"
/db_xref="taxon:4081"
/clone="cLEM26C18"
/clone_lib="tomato nutrient deficient roots"
/tissue_type="Roots"
/dev_stage="5-6 weeks old"
/lab_host="SOLR"
/note="Vector: pBluescriptSMKcladapT; Site_1: 5' EcoRI;
Site_2: 3' XhoI; Roots were harvested from plants grown under the following deficiencies/stresses: 10 mM Al, Zn, P, K, Fe, N. mRNA was isolated from individual treatments. Proportional aliquots of mRNA of each treatment were mixed and used for library construction."

BASE COUNT 149 a 115 c 117 g 115 t
ORIGIN

Query Match 1.2%; Score 20; DB 12; Length 496;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAAGATTCAGCTCCTGCTT 1590
|||||
Db 469 AGAAGATTCAGCTCCTGCTT 450

RESULT 34
AM651372/c

DEFINITION AM651372 498 bp mRNA linear EST 18-MAY-2001
EST329826 tomato germinating seedlings, TAMO Lycopersicon
esculentum cDNA clone CLE116M22 5', mRNA sequence.

ACCESSION AM651372
VERSION AM651372.1 GI:7412610

KEYWORDS EST.
SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 498)
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 498)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S. Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Niernon,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
S.D.
TITLE Generation of ESTs from germinating tomato seed
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1. 498
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEF16M2"
/clone_1lb="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."

BASE COUNT 149 a 112 c 123 g 114 t
ORIGIN

Query Match 1.2% Score 20; DB 10; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAAGATTACGCTCTGCTT 1590
|||||
Db 493 AGAAGATTACGCTCTGCTT 474

RESULT 35
BM404442/c 500 bp mRNA linear EST 22-JAN-2002
LOCUS EST578769 potato roots Solanum tuberosum cDNA cPRO20D24 5'
DEFINITION end, mRNA sequence.
ACCESSION BM404442
VERSION BM404442.1 GI:18255884
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 500)
van der Hoeven,R., Sun,H., Karamcheva,S.A., Tsai,J., Van Aken,S.,
Uterback,T., Chiening,A., Bougri,O., Buell,C.R., Konning,C.,
Tanksley,S. and Baker,B.
TITLE Generation of ESTs from potato roots
JOURNAL Unpublished (2001)
COMMENT Contact: Research Genetics, Libraries Division
Tel: 1-800-711-6195
Email: cdna@resgen.com
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: T3.
Location/Qualifiers
1. 500
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cPRO20D24"
/clone_1lb="potato roots"
/tissue_type="roots"
/dev_stage="in vitro grown stem cuttings"

FEATURES
source

/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University, Tanksley lab;
sequencing: The Institute for Genomic Research. Roots were
isolated from in vitro grown stem cuttings on CM medium.
Roots were isolated two weeks after placing the stem
cuttings from in vitro grown plants on medium."

BASE COUNT 142 a 116 c 129 g 113 t
ORIGIN

Query Match 1.2% Score 20; DB 13; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAAGATTACGCTCTGCTT 1590
|||||
Db 493 AGAAGATTACGCTCTGCTT 474

RESULT 36
AM221154/c 514 bp mRNA linear EST 18-MAY-2001
LOCUS EST297623 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone CLEF13A10, mRNA sequence.
ACCESSION AM221154
VERSION AM221154.1 GI:6532838
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 514)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ann,S., Ronning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1. 514
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEF13A10"
/clone_1lb="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"

BASE COUNT 158 a 118 c 122 g 116 t
ORIGIN

Query Match 1.2% Score 20; DB 10; Length 514;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAAGATTACGCTCTGCTT 1590
|||||
Db 503 AGAAGATTACGCTCTGCTT 484

RESULT 37				
LOCUS	BB284364			
DEFINITION	BB284364 RIKEN full-length enriched, adult retina musculus cDNA.			
ACCESSION	BB284364			
VERSION	BB284364.1			
KEYWORDS	GI:8984813			
ORGANISM	Mus musculus			
REFERENCE	house mouse.			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
TITLE	BB284364 mRNA linear EST 01-AUG-2000			
JOURNAL	BB284364 RIKEN full-length enriched, adult retina musculus cDNA.			
COMMENT	BB284364.1 GI:8984813			
	URL: genome.res.gsc.riken.go.jp,			
	Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
	The Institute of Physical and Chemical Research (RIKEN)			
	1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan			
	Fax: 81-45-503-9226			
	Email: genome.res.gsc.riken.go.jp,			
	URL:http://genome.gsc.riken.go.jp/			
	Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
	Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)			
	Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.			
	Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)			
	Carninci,P. and Hayashizaki,Y.			
	High efficiency full-length cDNA cloning. Methods Enzymol. 303, 15-44 (1999)			
	Please visit our web site (http://genome.rtc.riken.go.jp) for further details.			
FEATURES	Location/Qualifiers			
SOURCE	1..546			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/clone="A930105K03"			
	/clone_lib="RIKEN full-length enriched, adult retina"			
	/tissue_type="retina"			
	/dev_stage="adult"			
	/lab_host="DH10B"			
	/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGAGATCCAGACGCTCTTTTGTGGTTTTVVN 3'], cDNA was transcribed by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of			

```

sequence [5'GAGAGAGAGATTCCTGCAATTATTAATTAATCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. ~Retina RNA was provided by Stefano Gustincich,
Department of Neurobiology, Harvard Medical School, 220
Longwood Ave., Boston, MA02115, USA, whose assistance we
gratefully acknowledge."
BASE COUNT      180 a      112 c      103 g      151 t
ORIGIN

Query Match      1.2%: Score 20; DB 10; Length 546;
Best Local Similarity 100.0%; Pred. NO. 1.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1333 ACCAAGACCAAAAGTCAAA 1352
|||||
Db      328 ACCAAGACCAAAAGTCAAA 347

RESULT 38
LOCUS      A0121410      548 bp      DNA      linear      GSS 22-SEP-1998
DEFINITION HS.3087_B2_B12_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3087 Col=24 Row=D, DNA sequence.
ACCESSION  A0121410
VERSION     A0121410.1 GI:3498576
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 548)
AUTHORS    Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
High throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3087 row: D column: 24
Class: BAC ends
High quality sequence stop: 548.
Location/Qualifiers
1. 548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3087 Col=24 Row=D"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC1; BAC Clones in
E-Coli DH10B"
BASE COUNT      191 a      93 c      109 g      149 t      6 others
ORIGIN

Query Match      1.2%: Score 20; DB 17; Length 548;
Best Local Similarity 100.0%; Pred. NO. 1.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 AATATAAAACTATCAGAA 21
|||||
Db      167 AATATAAAACTATCAGAA 186

RESULT 39
Q0119131

```

LOCUS BQ119131 558 bp mRNA linear EST 17-APR-2002
 DEFINITION EST604707 mixed potato tissues Solanum tuberosum cDNA clone STMEH45
 3' end, mRNA sequence.
 ACCESSION BQ119131
 VERSION BQ119131.1 GI:20171093
 KEYWORDS EST.
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; easterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 558)
 Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,
 Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and
 Karamycheva, S.A.
 Generation of a set of potato cDNA clones for microarray analyses
 Unpublished (2002)
 CONTACT Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato@tigr.org
 This clone is available through the Research Genetics, contact the
 Research Genetics for further information 1-800-711-6195 or
 cdna@resgen.com
 Seq primer: 77.
 FEATURES
 source
 1..558
 /organism="Solanum tuberosum"
 /cultivar="Kennebec or Binje"
 /db_xref="taxon:4113"
 /clone="STMEH45"
 /clone_lib="mixed potato tissues"
 /tissue_type="mixed tissues"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
 XhoI; Suppl: Combination of untreated and phytophthora
 infestans-treated libraries of stolons, leaves, leaflets,
 axillary buds of stem explants, petioles, germinating eyes
 , tubers, or roots."
 BASE COUNT 170 a 120 c 91 g 177 t
 ORIGIN
 Query Match 1.2%; Score 20; DB 14; Length 558;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1270 TTTGCAATCCGCAAAATAC 1289
 ||||||||||||||||||
 Db 18 TTTGCAATCCGCAAAATAC 37
 RESULT 40
 BE459902/c 559 bp mRNA linear EST 18-MAY-2001
 LOCUS
 DEFINITION EST415194 tomato developing/immature green fruit Lycopersicon
 esculentum cDNA clone CLEM8120, mRNA sequence.
 ACCESSION BE459902
 VERSION BE459902.1 GI:9504204
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; easterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1 (bases 1 to 559)
 Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,
 Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Romling, C.M.,
 Niernan, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley,
 S.D.
 Generation of ESTs from tomato fruit tissue, immature green
 Unpublished (2000)
 CONTACT CUGI

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 FEATURES
 source
 Location/Qualifiers
 1..559
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CLEM8120"
 /clone_lib="tomato developing/immature green fruit"
 /tissue_type="fruit"
 /dev_stage="immature green (5-35 days post-anthesis)"
 /lab_host="SOLR"
 /note="Vector: pBluescriptSMCudapt; Site.1: EcoRI;
 Site.2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
 harvested at 7 day intervals through 35 dpa. Equal masses
 of tissue from each stage were combined (including seeds
 and locules) prior to mRNA isolation."
 BASE COUNT 174 a 127 c 129 g 129 t
 ORIGIN
 Query Match 1.2%; Score 20; DB 10; Length 559;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1571 AGAGATTTCAGCTCCTGCTT 1590
 ||||||||||||||||||
 Db 504 AGAGATTTCAGCTCCTGCTT 485
 Search completed: January 21, 2003, 21:43:48
 Job time : 1696 secs

THIS PAGE BLANK (USPTO)

QY 1561 AAGTGCATCAAGATTCAGTCTGCTTCTATAGAGTTGCTTCTTACTTCTT 1620
 |||
 Db 405880 AAGTGCATCAAGATTCAGTCTGCTTCTTCTATAGAGTTGCTTCTTACTTCTT 405939
 |||
 QY 1621 GATCCCTACCTGCTTTT 1637
 |||
 Db 405940 GATCCCTACCTGCTTTT 405956
 |||

RESULT 3
 AACB1914/c
 ID AACB1914 standard; DNA: 273254 BP.

AC AACB1914:
 XX
 XX 27-FEB-2001 (first entry)
 DE Chlamydia pneumoniae genome DNA.
 XX
 KM Genome: diagnosis: vaccine: ds.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN W0200027994-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 12-NOV-1999; 99W0-U526923.
 XX
 PR 12-NOV-1998; 98U5-0108279.
 PR 08-APR-1999; 99U5-0128606.
 XX
 PA (RECC) UNIV CALIFORNIA.
 XX
 PI Stephens R, Mitchell W, Kalman S, Davis R;
 XX
 DR WPI: 2000-376516/32.
 XX
 PT Isolated nucleic acid for use in diagnostic and analytical methods
 PT encodes genomic sequence of Chlamydia pneumoniae -
 XX
 PS Claim 2: Page 128-320; 320p; English.

CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia
 CC pneumoniae protein (P1), given in the specific method. The isolated nucleic
 CC acid is useful for diagnostic and analytical methods, such as,
 CC hybridization-based assays or amplification-based assays. The protein may
 CC be used for diagnostic purposes, for their enzymatic or structural
 CC activity, or as a vaccine. The invention also describes (1) a probe
 CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
 CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
 CC cassette comprising N1 under the transcriptional regulation of a
 CC transcriptional initiation region functional in an expression host, and a
 CC cassette of (3) as part of an extrachromosomal element or integrated into
 CC the genome of a host cell as a result of induction of the expression
 CC cassette into the host cell, and the cellular progeny of the host cell;
 CC (5) a method for producing a P1 comprising growing a cell of (4) where
 CC the protein is expressed and isolating the protein free of other
 CC proteins; (6) a purified polypeptide composition comprising at least 50
 CC weight % of P1; and (7) a monoclonal antibody binding specifically to the
 CC peptide of (6).
 CC
 XX
 XX Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Query Match 99.7%; Score 1632.2; DB 21; Length 273254;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATAAATAAATCATGATAGAAATTAAGATTTTCAGAGGTAATATGACAAAAA 60
 |||
 Db 211390 GAATAAATAAATCATGATAGAAATTAAGATTTTCAGAGGTAATATGACAAAAA 212411
 |||

QY 61 CCGAAGAAAAACCTTTTGGAAAAATGCGCTTCTTGTGGCCGATACATCTACAGAGC 120
 |||
 Db 212410 CCGAAGAAAAACCTTTTGGAAAAATGCGCTTCTTGTGGCCGATACATCTACAGAGC 212351
 |||
 QY 121 TAAAGAAAGTTTGGCCAAATGTTCTTAATGTTCTGTATTTACATTTAATACGGTGT 180
 |||
 Db 212350 TAAAGAAAGTTTGGCCAAATGTTCTTAATGTTCTGTATTTAATACATACGGTGT 212291
 |||
 QY 181 TACGGATACAAAAAGACACTTATGTGGAGACTCGTGTCTGTGCGAGAGCAATAC 240
 |||
 Db 212290 TACGGATACAAAAAGACACTTATGTGGAGACTCGTGTCTGTGCGAGAGCAATAC 212231
 |||
 QY 241 CTTTCATCAAGTTTGGCTTGTGCTCCCGCTGCTATATCTTATGCTTATTTATGCAA 300
 |||
 Db 212230 CTTTCATCAAGTTTGGCTTGTGCTCCCGCTGCTATATCTTATGCTTATTTATGCAA 212171
 |||
 QY 301 AGCTAAGTAATTTTATAGTAAGCAGGCTTATTTATGCAAGTGGGAGAGCCCTTTTAA 360
 |||
 Db 212170 AGCTAAGTAATTTTATAGTAAGCAGGCTTATTTATGCAAGTGGGAGAGCCCTTTTAA 212111
 |||
 QY 361 TTTTCTTTGGCCCTGTCGCGAGCTATTTATCCGCTACGCGATGTTTACATCTACAG 420
 |||
 Db 212110 TTTTCTTTGGCCCTGTCGCGAGCTATTTATCCGCTACGCGATGTTTACATCTACAG 212051
 |||
 QY 421 AATTGCTGACCGCTTTACAGGCCATCTACCTCCAGAGATTGCTAGAGACGCTGCTATCT 480
 |||
 Db 212050 AATTGCTGACCGCTTTACAGGCCATCTACCTCCAGAGATTGCTAGAGACGCTGCTATCT 211991
 |||
 QY 481 TAAAGAACTGAGACATTTGCTGCAATTTATGTAAGTCTGTAAGTGGGAGAGCGTATGC 540
 |||
 Db 211990 TAAAGAACTGAGACATTTGCTGCAATTTATGTAAGTCTGTAAGTGGGAGAGCGTATGC 211931
 |||
 QY 541 TATCTCAATGTTCTGGGGATTTGCTAATGAATTCAAAAATCCAGAGCAACCGTT 600
 |||
 Db 211930 TATCTCAATGTTCTGGGGATTTGCTAATGAATTCAAAAATCCAGAGCAACCGTT 211871
 |||
 QY 601 TCTAGCTCTTTTGGATTCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTTG 660
 |||
 Db 211870 TCTAGCTCTTTTGGATTCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTTG 211811
 |||
 QY 661 TTTGGGCTCAAGTTGAGAGCTCCGTTTCTGAAGGTATGATCCTTGGGAAATTTCTT 720
 |||
 Db 211810 TTTGGGCTCAAGTTGAGAGCTCCGTTTCTGAAGGTATGATCCTTGGGAAATTTCTT 211751
 |||
 QY 721 TACGCTTTTGAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 |||
 Db 211750 TACGCTTTTGAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 211691
 |||
 QY 781 GGATCAATAGAAAGCTATTGACCGATCTCGCTTCTAATATCCAGAAAGAAATGCAAAAG 840
 |||
 Db 211690 GGATCAATAGAAAGCTATTGACCGATCTCGCTTCTAATATCCAGAAAGAAATGCAAAAG 211631
 |||
 QY 841 GGAAAAAGGCTGTAACCTTAAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 900
 |||
 Db 211630 GGAAAAAGGCTGTAACCTTAAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 211571
 |||
 QY 901 CTCCTTAATATCTTTTATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 |||
 Db 211570 CTCCTTAATATCTTTTATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 211511
 |||
 QY 961 TCGAAGTGAAGTGAAGAACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1020
 |||
 Db 211510 TCGAAGTGAAGTGAAGAACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 211451
 |||
 QY 1021 AGTTCATGGGGAACCTTCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1080
 |||
 Db 211450 AGTTCATGGGGAACCTTCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 211391
 |||
 QY 1081 TTGGTGTACCTCAATGCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 |||
 Db 211390 TTGGTGTACCTCAATGCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 211331
 |||
 QY 1141 TGGTTCCTCAACAGGATGCTTTTCTGCTCTTGTATATCTTTTGAAGAACCAAGCTTGTG 1200

THIS PAGE BLANK (USPTO)